

THE CPS UTILITIES USER'S MANUAL

Unicon Research Corporation
1640 Fifth Street, Suite 100
Santa Monica, CA 90401

September 2005

Copyright 1994-2005 by Unicon Research Corporation
All Rights Reserved

Suggested citation formats

Unicon Research Corporation provides this manual, as well as the accompanying Data Dictionary and compact discs, on an "as is" basis. These products bear no warranty of any kind, either expressed or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Unicon Research Corporation may make improvements and/or changes in any of these products at any time and without notice.

The original collector of the data, Unicon Research Corporation, and the relevant funding agency bear no responsibility for the uses of this collection or for the interpretation of inferences based upon such uses.

Publications based on the CPS Utilities data collection should acknowledge those sources by means of bibliographic citations. To ensure that such source attributions are captured for economic and social science bibliographic utilities, citations must appear in footnotes or in the reference section of publications.

The **abbreviated bibliographic citation** for these data is:

"Current Population Survey, [appropriate file and year(s)], CPS Utilities, Unicon Research Corporation, www.unicon.com" where the file would be March, Outgoing Rotations, Tobacco Use, etc.

It is suggested that professional publications use the following, more detailed, citations:

Data Files

Workplace Topics I (January/February/Unemployment Compensation) data files:

CURRENT POPULATION SURVEYS, JANUARY 1973, 1975-1978, 1980-1984, 1986-1988, 1990-1993, 1996, 1999, 2002, 2004; FEBRUARY 1984, 1990, 1994-2002 MAY 1989, AUGUST 1989, and NOVEMBER 1989 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1975-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Annual Social and Economic (March) data files:

CURRENT POPULATION SURVEYS, MARCH 1962-2005 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1962-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Workplace Topic II (May/September) data files:

CURRENT POPULATION SURVEYS, MAY 1969-1989, 1991, 1993, and 1995-1997, 1999-2001, SEPTEMBER 2002 and 2003, DECEMBER 1989 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1969-2004. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Marriage and Fertility (June) data files:

CURRENT POPULATION SURVEYS, JUNE 1971, 1973-1977, 1979-1988, 1990-1992, 1994-1995, 1998, 2000, 2002, and 2004 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1971-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Education and School Enrollment (October) data files:

CURRENT POPULATION SURVEYS, OCTOBER 1968-2003: SCHOOL ENROLLMENT [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1968-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Voting and Voter Registration (November) data files:

CURRENT POPULATION SURVEYS, NOVEMBER 1972, 1974, 1976, 1978-1980, 1982-1994, 1996, 1998, 2000, 2002, and 2004 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1975-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Agricultural Workers data files:

CURRENT POPULATION SURVEYS, DECEMBER 1977, 1979, 1981, 1983, 1985, and 1987 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1977-1987. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Child Support and Alimony data files:

CURRENT POPULATION SURVEYS, MARCH/APRIL MATCH 1979, 1982, 1984, 1986, 1988, 1990, 1992, 1994, and 1996 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1979-1996. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Food Security data files:

CURRENT POPULATION SURVEYS, APRIL 1976, AUGUST 1977, APRIL 1995, SEPTEMBER 1996, APRIL 1997, AUGUST 1998, APRIL 1999, SEPTEMBER 2000, APRIL 2001, DECEMBER 2001, and DECEMBER 2002 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1976-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Immunization data files:

CURRENT POPULATION SURVEYS, SEPTEMBER 1978-1985 and 1989 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1978-1989. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Internet and Computer Usage data files:

CURRENT POPULATION SURVEYS, OCTOBER 1984, 1989, 1993, 1997, NOVEMBER 1994, DECEMBER 1998, AUGUST 2000, and SEPTEMBER 2001 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1984-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Tobacco Usage data files:

CURRENT POPULATION SURVEYS, SEPTEMBER 1985, 1989, 1992, 1995, 1998; JANUARY 1993, 1996, 1999, 2000; MAY 1993, 1996, 1999, 2000; JUNE 2001, NOVEMBER 2001, and FEBRUARY 2002 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1985-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Veteran data files:

CURRENT POPULATION SURVEYS, APRIL 1985; NOVEMBER 1987; SEPTEMBER 1989, 1991, 1993, 1997, 1999; and AUGUST 1995 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1985-1999. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Outgoing Rotations Earner Study data files:

CURRENT POPULATION SURVEYS, 1979-2004 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1979-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Monthly Basic data files:

CURRENT POPULATION SURVEYS, JANUARY-DECEMBER 1976-2005 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 1976-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Outgoing Rotations Monthly Earner Study data files:

CURRENT POPULATION SURVEYS, JANUARY-DECEMBER 2005 [machine-readable data files]/conducted by the Bureau of the Census for the Bureau of Labor Statistics. Washington: Bureau of the Census [producer and distributor], 2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Data Dictionaries

Workplace Topics I (January/February/Unemployment Compensation) data dictionary:

CURRENT POPULATION SURVEYS, JANUARY 1973, 1975-1978, 1980-1984, 1986-1988, 1990-1993, 1996, 1999, 2002, 2004; FEBRUARY 1984, 1990, 1994-2002 MAY 1989, AUGUST 1989, and NOVEMBER 1989 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1973-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Annual Social and Economic (March) data dictionary:

CURRENT POPULATION SURVEYS, MARCH 1962-2005 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1962-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Workplace Topics II (May/September) data dictionary:

CURRENT POPULATION SURVEYS, MAY 1969-1989, 1991, 1993, and 1995-1997, 1999-2001, SEPTEMBER 2002 and 2003, DECEMBER 1989 TECHNICAL DOCUMENTATION/ prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1969-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Marriage and Fertility (June) data dictionary:

CURRENT POPULATION SURVEYS, JUNE 1971, 1973-1977, 1979-1988, 1990-1992, 1994-1995, 1998, 2000, 2002, and 2004 TECHNICAL DOCUMENTATION/ prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1971-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Education and School Enrollment (October) data dictionary:

CURRENT POPULATION SURVEYS, OCTOBER 1968-2003, TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1968-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

November Voting and Voter Registration data dictionary:

CURRENT POPULATION SURVEYS, NOVEMBER 1972, 1974, 1976, 1978-1980, 1982-1994, 1996, 1998, 2000, 2002, and 2004, TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1978-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Agricultural Workers data dictionary:

CURRENT POPULATION SURVEYS, DECEMBER 1977, 1979, 1981, 1983, 1985, and 1987 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1977-1987. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Child Support and Alimony data dictionary:

CURRENT POPULATION SURVEYS, MARCH/APRIL MATCH 1979, 1982, 1984, 1986, 1988, 1990, 1992, 1994, and 1996 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1979-1996. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Food Security data dictionary:

CURRENT POPULATION SURVEYS, APRIL 1976, AUGUST 1977, APRIL 1995, SEPTEMBER 1996, APRIL 1997, AUGUST 1998, APRIL 1999, SEPTEMBER 2000, APRIL 2001, DECEMBER 2001, and DECEMBER 2002 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1976-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Immunization data dictionary:

CURRENT POPULATION SURVEYS, SEPTEMBER 1978-1985 and 1989 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1978-1989. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Internet and Computer Usage data dictionary:

CURRENT POPULATION SURVEYS, OCTOBER 1984, 1989, 1993, 1997, NOVEMBER 1994, DECEMBER 1998, AUGUST 2000, and SEPTEMBER 2001 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1984-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Tobacco Usage data dictionary:

CURRENT POPULATION SURVEYS, SEPTEMBER 1985, 1989, 1992, 1995, 1998; JANUARY 1993, 1996, 1999, 2000; MAY 1993, 1996, 1999, 2000; JUNE 2001, NOVEMBER 2001, and FEBRUARY 2002 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1985-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Veteran data dictionary:

CURRENT POPULATION SURVEYS, APRIL 1985; NOVEMBER 1987; SEPTEMBER 1989, 1991, 1993, 1997, 1999; and AUGUST 1995 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1985-1999. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Outgoing Rotations Earner Study data dictionary:

CURRENT POPULATION SURVEYS, OUTGOING ROTATIONS 1979-2004 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1979-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Monthly Basic data dictionary:

CURRENT POPULATION SURVEYS, JANUARY-DECEMBER 1976-2005 forward TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 1976-2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

Outgoing Rotations Monthly Earner Study data dictionary:

CURRENT POPULATION SURVEYS, JANUARY-DECEMBER 2005 TECHNICAL DOCUMENTATION/prepared by Administrative and Customer Services Division, Microdata Access Branch, Bureau of the Census. Washington: The Bureau [producer and distributor], 2005. Santa Monica, CA: Unicon Research Corporation [producer and distributor of CPS Utilities], 2005.

THE CPS UTILITIES USER'S MANUAL

Table of Contents

Suggested citation formats	2
Table of Contents	8
1. ACKNOWLEDGMENTS	9
2. INTRODUCTION AND BACKGROUND	9
3. CONTENTS OF THE DATA DICTIONARY AND THE COMPACT DISC	10
4. INSTALLATION	12
4.1. SOFTWARE INSTALLATION	14
4.2. DOCUMENTATION INSTALLATION	15
4.3. HELPER APPLICATIONS	16
5. USING CPS UTILITIES	17
5.1 SEARCH UTILITY	20
5.1.1 WORD SEARCH	21
5.1.2 STRING SEARCH	22
5.1.3 SEARCH RESULTS	23
5.1.4. SAMPLE OUTPUT FILES: SEARCH UTILITY	23
5.2. DOCUMENTATION UTILITY	28
5.2.1. SAMPLE OUTPUT: DOCUMENTATION UTILITY	29
5.3 EXTRACT UTILITY	30
5.3.1 USING THE EXTRACT UTILITY	32
5.3.2 MULTITASKING	37
5.3.3 REQUEST FILES	38
5.3.4 SAMPLE OUTPUT FILES: EXTRACT UTILITY	43
5.3.5 RECODES	45
6. IMPORTANT INFORMATION AND TIPS FOR USERS	48
6.1 STRUCTURE OF THE CPS DATA FILES FOR MARCH AND CSA	48
6.2 STRUCTURE OF THE CPS DATA FILES FOR OTHER THAN MARCH AND CSA	48
6.3 MISCELLANEOUS TIPS ON USING THE UTILITIES	49
7. FUTURE ENHANCEMENTS TO THE SOFTWARE	50

1. ACKNOWLEDGMENTS

This product was funded in part by grants from the National Institute on Aging, the National Library of Medicine, the National Institute of Child Health and Human Development, and the U.S. Census Bureau. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding institutions.

With the 1997 release of the data, Unicon replaced the 1964-1967 Mare-Winship files with the original census files for 1962-1967. It was believed that these files were lost as neither the Census Bureau nor the National Archives had them. However the University of Wisconsin contacted us and volunteered to send us copies of these early files. We are indebted to them for their contribution of the data and documentation. They can be contacted at Data and Program Library Service, 3313 Social Science Building, 1180 Observatory Drive, Madison, WI 53706 (<http://dpls.dacc.wisc.edu>), (608) 262-7962. They were unable to locate the 1965 data file. However, a plea to our users led another user to contribute her copy of the 1965 data.

We would like to thank and acknowledge these people in aiding our efforts to preserve the original data. If you know of the existence of any earlier data files, please contact us. Unicon's goal is to archive and provide access to as many CPS files as can be located.

2. INTRODUCTION AND BACKGROUND

The Current Population Surveys (CPS) are administered monthly by the U.S. Bureau of the Census to over 65,000 households. These surveys gather a wealth of information on the education, labor force status, demographics, and other aspects of the U.S. population. The CPS is widely used by demographers, economists, sociologists, and others interested in population and demographic issues. In addition, it serves an important role as the basis upon which important federal statistics, like unemployment rates, are calculated.

Each month's CPS has a different emphasis. The October surveys, for example, focus on school enrollment while the June surveys feature marriage and fertility issues. The March surveys are especially important, for they contain the Annual Demographic File and the Income Supplement, which reports in detail the income-related aspects of U.S. households. Currently, the January data is available from 1973, the February data from 1984, the March data from 1962, the May data from 1969, the June data from 1971, the October data from 1968, the November data from 1972, the Agricultural Worker data from 1977 to 1985, the Immunization data from 1985 to 1989, the Internet and Computer Usage data from October 1984, the Child Support and Alimony data from March/April 1979, the Food Security data from April 1976, the Tobacco Usage data from September 1992, the Veteran data from 1985, the Outgoing Rotations (Earnings) data from 1979, the Monthly Outgoing Rotations data for the current year, and the Monthly Basic data from 1976 are available. The Census Bureau produces the raw data provided with the CPS Utilities. New files are added to the CPS series as they become available from the Census. Note that not all files are released annually. See our website, <http://www.unicon.com>, for a complete list of years available for each survey.

Despite their importance to the research community, the CPS files are inconvenient to use in several ways, particularly for the novice but even for those experienced in the use of these data. Problems are especially acute for those attempting to form a time series by piecing together surveys from many different years. Variables change location and length over time, requiring several different program formats to obtain a given set of variables across many years. Old variables are dropped and new ones added. Codings change from time to time, as do the questions from which the variables are derived. Moreover, these changes in questionnaire content are often subtle. The values at which monetary variables are top-coded (i.e., the unbounded top range of values, for instance 50+) vary over time, often in ways not clearly spelled out in the documentation supplied with the surveys.

Even obtaining data from a single year is not without its pitfalls. The Census-supplied documentation is sometimes incomplete and difficult to interpret, particularly for the older surveys. Determining the universe of respondents for particular questions is frequently not straightforward, requiring researchers to trace through skip patterns on questionnaires. Even the act of finding all variables on a specific topic, determining their coding, and ascertaining the context in which the appropriate questions were asked, can itself be a cumbersome process that requires a time-consuming manual search through the CPS documentation.

The CPS Utilities address all of these deficiencies and difficulties. Utilizing three basic functions that are described in detail below, the software permits the user to locate relevant variables with relative ease, to produce data files by simply naming variables and years, and to collect and view in compact form all coding, universe and questionnaire information for a variable across all survey years. Moreover, the system provides some information that is not available in the written documentation available from the Census, and it offers uniformly recoded versions of selected variables. Finally, the system facilitates improved accuracy of programming and speed of extraction.

CPS Utilities runs under Windows, providing the user with an intuitive visual interface. It includes a facility for creating and editing request files, and it displays the contents of the output files directly on the screen.

The CPS Utilities require a 386 (or higher) PC that is running under the Windows 95, 98, 2000, Me or NT operating systems, and is attached to a CD or DVD reader.

3. CONTENTS OF THE DATA DICTIONARY AND THE COMPACT DISC

Although the heart of the system is the software contained on the compact disc supplied with this manual, the voluminous CPS Data Dictionary is an important component as well, for it serves as a hardcopy codebook and supplementary source of information. After a brief introduction, the dictionary is divided into six sections:

- I. A listing of variables ordered according to the original CPS column locations. This feature may be useful for those users who are familiar with the original CPS documentation and who wish to find the variable names that have been assigned to the original CPS variables.

- II. A list of variables that have had their names changed in the CPS Utilities anytime since the original release of the software. We understand the importance in maintaining consistent names across time. However, as the software is supplemented and refined, we find that some of the names that were originally assigned to the variables are clumsy and not as user-friendly as we would like. This section lists the variables in alphabetical order by the original names and shows the new names that are currently assigned to the variables.
- III. An index of variables arranged under topic headings.
- IV. A codebook consisting of a listing of all variables, together with definitions, code values and meanings, top codes (if any), universe information, questionnaire items, and special notes. The contents of this codebook are also accessed by the system software, as described below.
- V. Appendices containing top codes, bottom codes, and missing value codes for income variables; codes for variables with values so numerous that they were deemed too long to include in the machine readable version of the codebook (including for example, 3-digit industry and occupation, state, and SMSA); and a few pages extracted from the original manuals that discuss geographic limitations of the file, weights, the earnings study, data preparation and value allocation, health insurance variables, and the process of matching files across years. These appendices are referenced in the documentation. If you use any of the utilities and are directed to an appendix, this is where you should look.
- VI. A section documenting changes made to the questionnaire (and hence the data) between 1993 and 1994.

Aside from the Data Dictionary and the User's Manual that you are now reading, the only other item that is part of the package is the compact disc. Although you do not need to be concerned with the structure of the CD or DVD in order to use the system successfully, it may be of interest to know its contents. The CD or DVD contains the following major directories or files:

- The directories "program files" and "System32" contain the installation files for the CPS Utilities software.
- On the March CD/DVD, the directory "cpsmar" contains the codebooks, catalogs, and tallies for the March CPS in a format that can be accessed by the system software. Discs for other survey months contain similarly named directories containing corresponding data.
- The directory "data" contains compressed CPS data in separate files by year.
- The directories "manual" (Data Dictionary and Glossary), "questionnaires" (Census Bureau questionnaires), and "Census docs" (Census Bureau technical documentation) contain information relevant to a specific survey in Adobe Portable Document Format.
- The directory "documentation" contains appendix files and other information relevant to all surveys in Adobe Portable Document Format.

Some of these directories and files are put on your machine during the course of installation, while others will continue to reside only on the disc, or, at your option, will be copied to your machine. In particular, the installation routine places the directory "cpsmar" (or "cpsout", etc.) and its subdirectories on your hard drive, as well as general documentation such as the appendices, "CPS User's Manual.pdf" and "TopicBySurvey.xls" in the directory "documentation". (Note that the installation routine makes these directories for you; do not attempt to do so yourself.) The directory "data" occupies so much space that users might not want it on their hard disk, and thus it generally remains only on the CD/DVD. Similarly, the directories "manual" and "questionnaires" are generally only on the CD/DVD. However, they may be easily copied to your machine. See Section 4.2. Documentation Installation below for details.

4. INSTALLATION

Installation of a new CD-ROM/DVD comprises both software and dictionary installation. Typically, the data files and most documentation files remain on the CD-ROM/DVD.

Beginning with version 5.3 of CPS Utilities software, some new features have been added, and this User Manual has been changed accordingly. We have changed from using a file named cpsutl.ini that is stored in the Windows directory to using the CURRENT_USER registry because several of our users have had difficulty getting the proper permissions to use the Windows directory in a networked environment. We have made copying data files and survey-specific documentation files easy to copy to your hard disk, should you choose to do so. Access to this documentation has been added as a 'General Documentation' button on the Entry screen. Saving request files to a different working directory has been made easier, as you can now change directories from the Utilities screen, in addition to using the Change Setup screen. For all surveys, a 'type' filter, describing the origin of the variable, has been added to display any combination of variables from the supplement questionnaire, basic questionnaire and those which are internally generated by the Census. For the March Annual Social and Economic Study and Child Support and Alimony surveys, a 'level' filter has been added to display any combination of household, family and person hierarchical level variables. We are also using InstallShield® for installations.

These new features are only available with version 5.3 and later, and if the dictionary files for the survey on a 5.3 CD have been updated using Reload from the Change Setup screen. As updates are released, the new features will become available for each survey.

We recognize that some users do not always update the software when they receive new CDs. To maintain compatibility with older software, the same information that has always been available will continue to be on the CDs/DVDs

However, for the 2004 Workplace Topics I (January/February), Earner Study (Outgoing Rotations), and both Monthly surveys (Basic and Earner Study, starting with May), we have made some changes that require upgrading to version 5.4 of CPS Utilities.

If the current version of the CPS Utilities software has already been installed, it does not need to be re-installed. To check the version number of the existing installation, click Help in the menu

bar of any screen, and then click About CPS Utilities. The version number is displayed in the top line of the text. If it is equal to or greater than the version number marked on the CD-ROM/DVD, the software installation may be skipped, but the new dictionary must be installed.

4.1. SOFTWARE INSTALLATION

If you are currently a CPS Utilities user and are installing version 5.3 or later for the first time, use the Add/Remove Programs applet in the Windows Control Panel to remove all previously installed versions. After installing it for the first time, InstallShield® will automatically remove the old version and install the new version for all subsequent updates. If you are not installing version 5.3, it is not necessary to uninstall prior versions, as obsolete files will be ignored. If you want to remove them from your disk anyway, use the Add/Remove Programs applet in the Windows Control Panel.

To install the CPS Utilities program files from the CD-ROM/DVD, place it into your disc reader. If your computer is set up to allow discs to "autorun", and if the current version of the CPS Utilities program files is not already installed, in a few seconds you will be asked whether you want to install them. Click Yes and follow the screen instructions.

Alternatively, you may install the program files from Windows Explorer. Find the disc reader drive letter and right-click the CD/DVD icon. Click Install from the options menu and follow the instructions.

A third option is to install the program files from Windows: Click Start, click Run, and in the Open box type: d:\setup (where d: is the drive letter of your disc reader). Click OK, and follow the screen instructions.

4.2. DOCUMENTATION INSTALLATION

When the software installation is complete, CPS Utilities will start, showing a Welcome screen (Figure 1). This screen allows you to specify a directory for your data dictionary and working files, as well as the drive letter of your disc reader. If you had previously installed CPS Utilities, this screen will be titled Update Files and your existing selections will be displayed automatically. Specifying the Data file directory is new with version 5.3, and will default to the CD/DVD drive and directory \data. Optionally, you may select to copy the data files and other survey-specific documentation files to your hard drive by clicking the appropriate buttons.

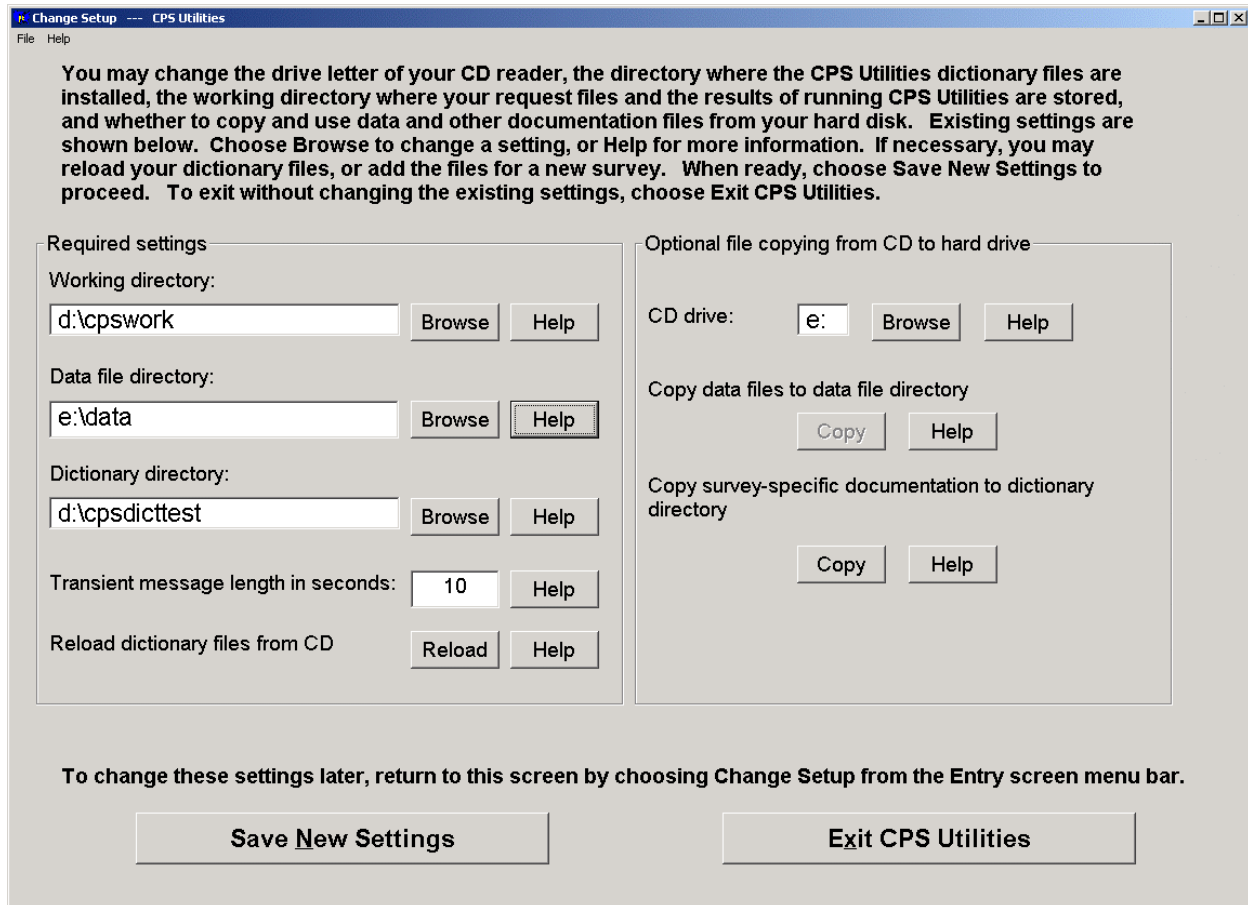


Figure 1. Welcome/Change Setup/Update Files screen

When you are ready, choose the Continue button. This will copy all the Dictionaries and other necessary files from the CD/DVD to your hard disk. The data files will stay on the disc, unless you copy them to your hard drive by specifying the data file directory on your hard drive and clicking the Copy button to the right. Likewise, the manual and questionnaire files will remain on the disc, unless you click the Copy button to the right to copy them to your dictionary directory. When you receive annual update CDs/DVDs, use the Reload button to update the dictionary files, the Copy buttons to update your data and documentation files, if desired, and the Save New Settings button to continue.

When the installation is complete, the Entry screen will show, and you will be ready to start using CPS Utilities. Once you have completed the installation, the Welcome screen will not show again unless you choose 'Change Setup' from the Entry screen menu bar.

If you had skipped the software installation because the current version was already installed, install the new dictionary files by choosing 'Change Setup', and then clicking 'Reload'. Optionally, update your data files and documentation using the 'Copy' buttons.

4.3. HELPER APPLICATIONS

To utilize all the features of CPS Utilities, three third-party "helper" applications must be present on the user's computer. Probably, they are already installed, but in case they are not, they may be downloaded free of charge from their publishers' web sites, using the links below.

The first is the Acrobat Reader mentioned above. Version 4.05 or later is required to view this User's Manual and the Data Dictionary .pdf files. If necessary, download the latest version from Adobe (<http://www.adobe.com/products/acrobat/readstep.html>).

The second is a subset of Microsoft® Internet Explorer. Version 4.0 or later is required to view the Documentation window on the Utilities screen. Internet Explorer does not need to be the default browser, or even present in full on the computer. The necessary files are included with Windows 98 or later, and Windows NT 4.0 or later. If you are using Windows 95, though, and Internet Explorer is not already installed, download it from Microsoft (<http://www.microsoft.com/downloads/>).

The third is a viewer for Microsoft® Excel, required to read the Topic By Survey file. If Excel is not installed on your machine, a viewer may be downloaded free of charge from Microsoft (<http://www.microsoft.com/office/000/viewers.asp>).

(If any of those links fails, the publisher has rearranged their site since we prepared this manual. Try the Adobe (<http://www.adobe.com/>) or Microsoft (<http://www.microsoft.com/>) home pages).

The file helpers.htm, in the root directory of the CD-ROM/DVD, can be read in any web browser and used to link to the respective download sites.

Follow the publishers' instructions to install the downloaded applications.

When you have finished, choose the Continue button. When the installation is complete, the Entry screen will show, and you will be ready to start using CPS Utilities. Once you have completed the installation, the Welcome screen will not show again unless you choose Change Setup from the Entry screen menu bar.

5. USING CPS UTILITIES

The SEARCH utility helps you locate all CPS variables related to a particular concept and displays them as a list. It also displays and produces a file containing the variable name, definition, coding, universe and questionnaire items for each such variable. The DOCUMENTATION utility produces that file for individual variables that you select by name. The EXTRACT utility produces a dataset containing variables that you specify, together with a documentation report on all of the variables in your data set, and program code files to read the dataset into any of three statistical analysis software packages (Stata, SAS, and SPSS).

In addition to providing all the raw variables, we also offer the facility to construct numerous recodes (primarily for the March series), which are distinguished from the raw variables by the presence of a leading underscore "_". These recoded variables are documented in the Data Dictionary along with the raw variables. Additional information on the nature and use of recodes is presented in Section 5.3.5.

The three CPS Utilities have been integrated into a single Utilities screen, which is accessed from the Entry screen (Figure 2). Select a survey, click "Start", and you are ready to go to work. For an overview of how CPS Utilities works, click "Introduction to CPS Utilities".

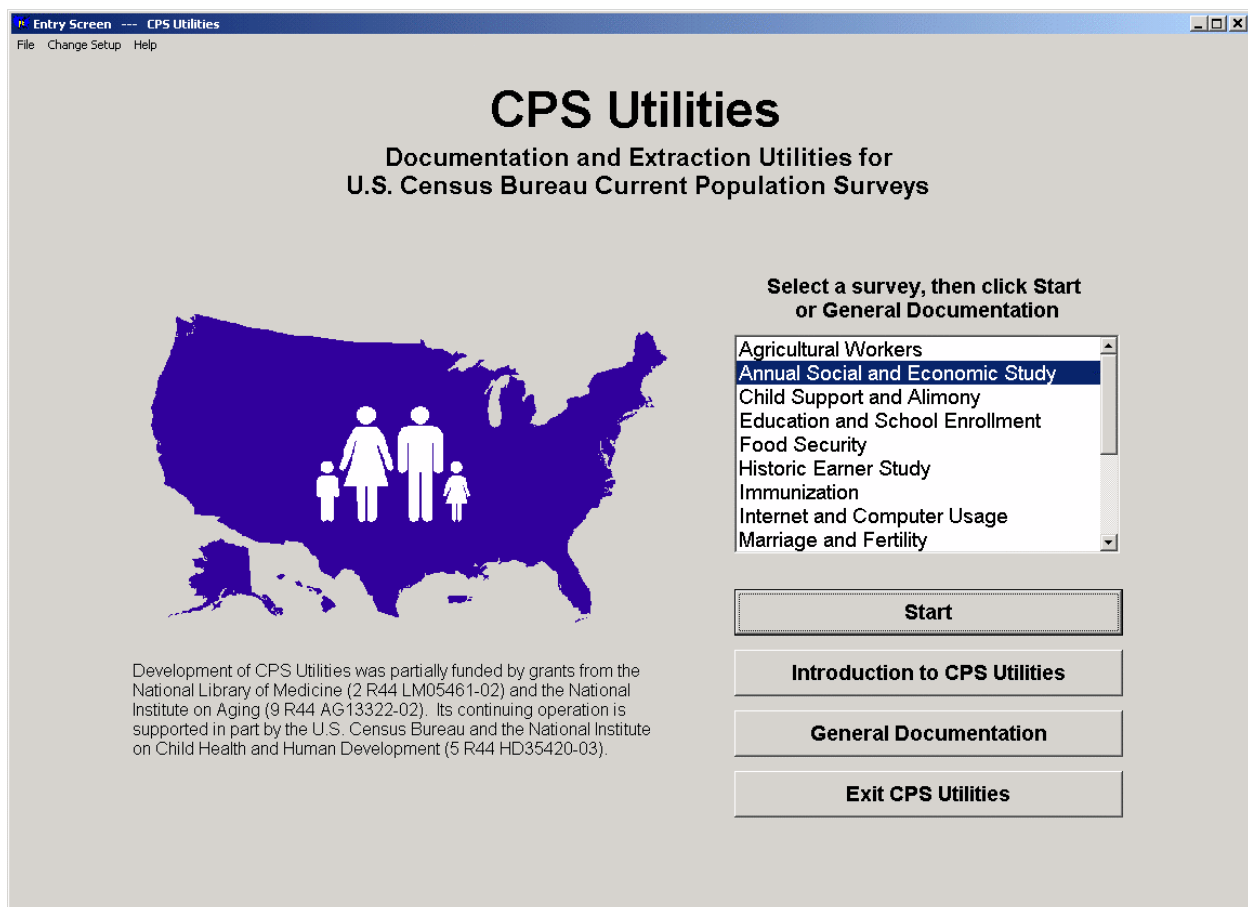


Figure 2. Entry screen

The "General Documentation" button gives you on-line access to all of the documentation available for all surveys, as well as for specific surveys (Figure 3). The Documentation Guide provides an overview of the documentation that is available on this screen. If any of the buttons are disabled, or 'grayed', it means that those files are not immediately available. You will be prompted to either load the appropriate installation CD/DVD for the selected survey and view them from the disc, or go to the Change Setup screen, load the CD/DVD and copy the survey-specific files to your hard drive. Older CDs may not have the questionnaire or glossary files on them.

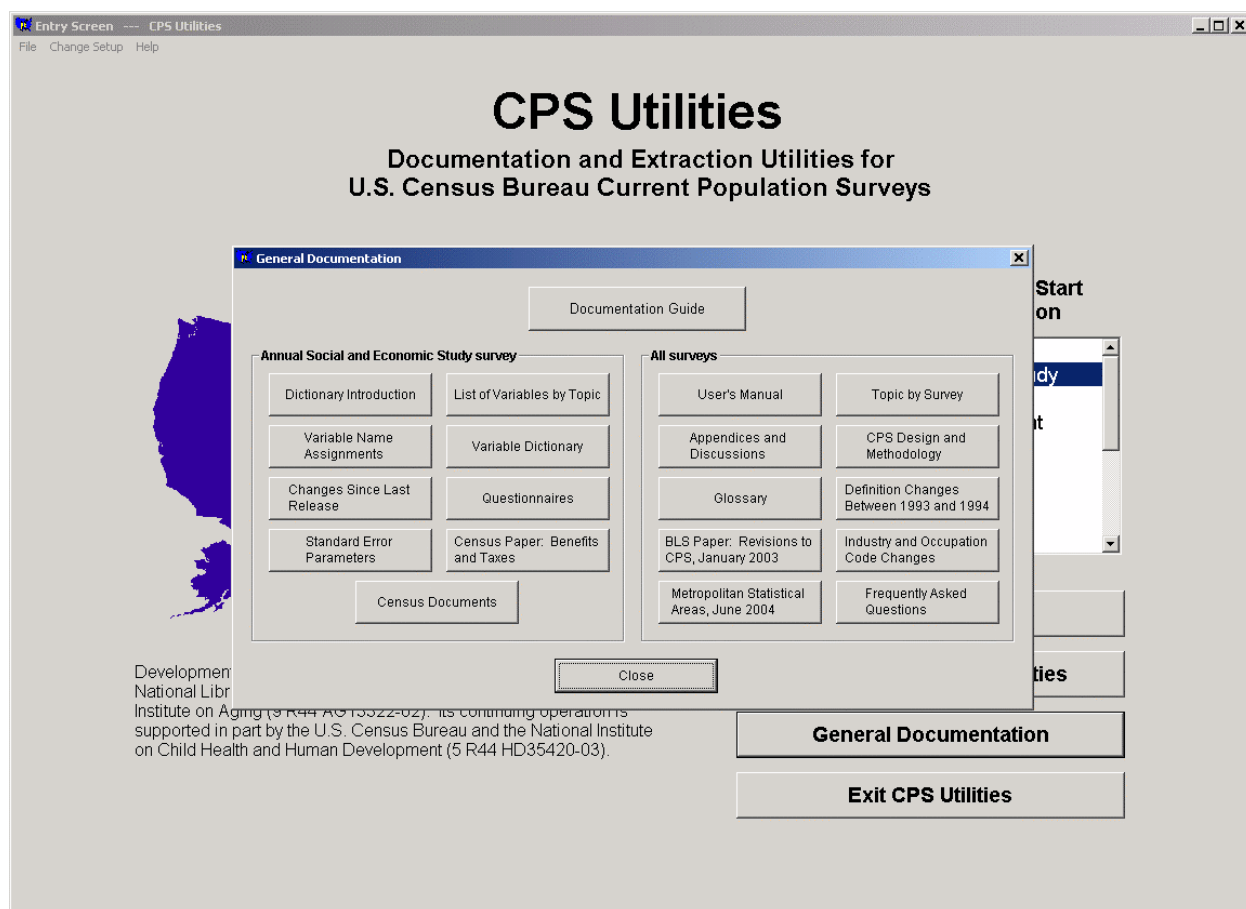


Figure 3. General Documentation screen

The Utilities screen (Figure 4) is the main gateway into CPS Utilities, allowing the user to run any of the three original utilities (SEARCH, DOCUMENTATION and EXTRACT). Two lists of variables are shown. The upper list lets you list variables alphabetically, by topic, by most recent search, optionally filtered by selected year(s). Two new filters also let you select type(s) and level(s), type being supplement, basic and internal, level being household, family and person. If the survey is not Annual Social and Economic (March) or Child Support and Alimony, only the type filter is available. Additionally, they may be displayed with titles included, or as just the variable names, eight columns across. The lower list is where the request file is prepared. The familiar SEARCH, DOCUMENTATION and EXTRACT buttons and functionality are still there, now all on one screen. "Extraction options" contains all of the report, program code, output file, labels, preview, conditional recoding and record limit options. Previously available on several screens, they are now conveniently located on several tabs on just one screen. Request file editing options are just a right-click away. Also new is the ability to change your working directory when you make a new request file without leaving the Utilities screen.

For more help, press F1 or choose Help from the menu bar.

The screenshot shows the 'Utilities Screen' window for the 'March CPS: Annual Social and Economic Study'. The window is divided into several sections:

- Top Section:** A title bar and a menu bar (File, Help). Below is a heading: "Use this section to find your Annual Social and Economic Study CPS variables, and add them to request files."
- Left Panel (List variables):** Contains radio buttons for "Alphabetically" (selected), "By topic", and "Search result". There are checkboxes for "Show titles" and "Filter by" (with sub-options for "Year(s)", "Type", and "Level"). Buttons include "Change topic", "Search", "Last search report", and "Documentation".
- Variable List:** A table of 1201 variables (all years, all types, all levels) displayed in 8 columns. Variables include aage, adepriv, afsval, ahealth, ahityp1, aincalm, aincrnt, aincs26, aincswc, etc.
- Bottom Section:**
 - Left Panel (Selected variable):** Includes an "Add all" button, a text field for "Type variable name or select from list above:", and an "Add" button.
 - Right Panel (To extract CPS data...):**
 - Request file name:** A list box showing "newout", "order", "out", "outrecodes", "outrecodes2", "recodes", "rectype", "spneth", "sttax", "test", "Change directory", "New request", and "Delete request".
 - Year(s):** A list box showing years from 2005 down to 1996.
 - Request file contents:** A text area showing "recnum", "age", "incwag", and "race".
 - Buttons:** "Extract", "Extraction options", "Last extraction report", "Help", and "Return to Entry Screen".

Figure 4. Utilities screen

5.1 SEARCH UTILITY

There are thousands of variables logged in the Data Dictionaries. More are being added each year. Searching through the possibilities to find a particular variable can proceed in several ways. If you are already familiar with the original CPS manuals provided by the Census Bureau and know which variables you are after by their original CPS names, you may want to use Section I of the Data Dictionary. There we list variables according to column location and year in the original CPS files, and it is easy to determine what names we have assigned to the variables. An alternative method is to search for variables by looking at what is available under general topics. Section III of the Data Dictionary facilitates this search by listing the variables grouped by topic.

Finally, using the CPS Utilities in a more substantive way, you may employ computer-assisted search through the SEARCH utilities. Upon choosing SEARCH on the Utilities screen, the Specify Search screen is displayed. Click on either the "Word Search" tab or the "String Search" tab, fill in the appropriate information, mark checkboxes as required, and click "Start".

5.1.1 WORD SEARCH

This utility provides a fast and flexible search capability utilizing an index of all the words in the data dictionary. It searches with case-insensitivity, allowing, for instance, both 'Spanish' and 'spanish' to be found with a single search specification.

Examples of search specifications are shown on the Word Search screen (Figure 5), enabling the user to easily specify any Boolean combination of search words. An asterisk may be used as a wildcard, to represent any ending. The search is conducted over variable names and definitions, coding, universe information, and questionnaire context for all variables in all of the CPS years selected. If a word that you supply is included within any element of an entry in the data dictionary, then the corresponding entry is included in the results. See Section 5.1.3 for details on how to use the results.

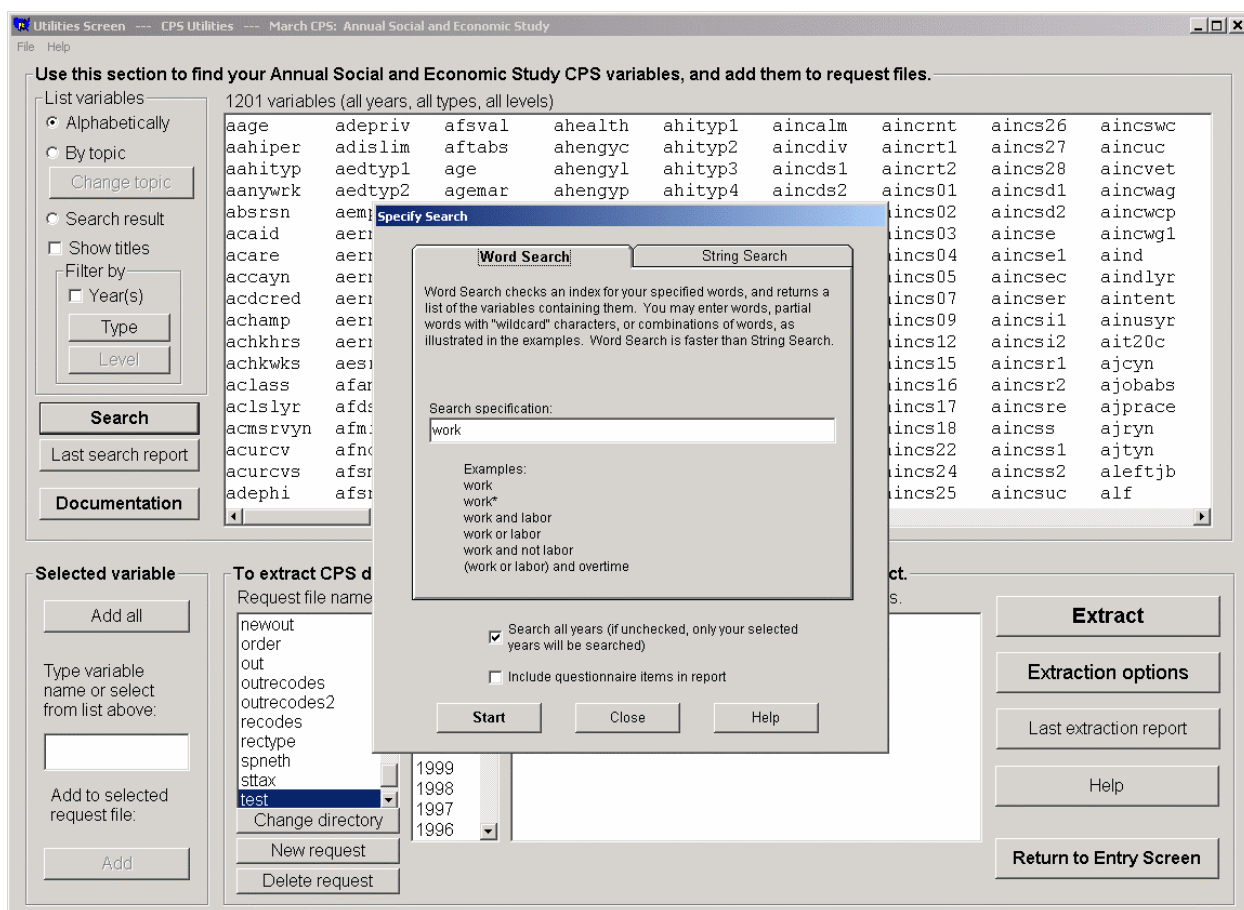


Figure 5. Specify Search screen for Word Search

5.1.2 STRING SEARCH

The "String Search" tab provides a different SEARCH utility to be used if required, for instance to perform a case-sensitive search, to search for strings containing multiple words separated by spaces, or to search for character strings embedded within words.

To use the String Search utility (Figure 6), enter one or more words or character strings that are related to the subject of interest in the boxes, starting from the top.

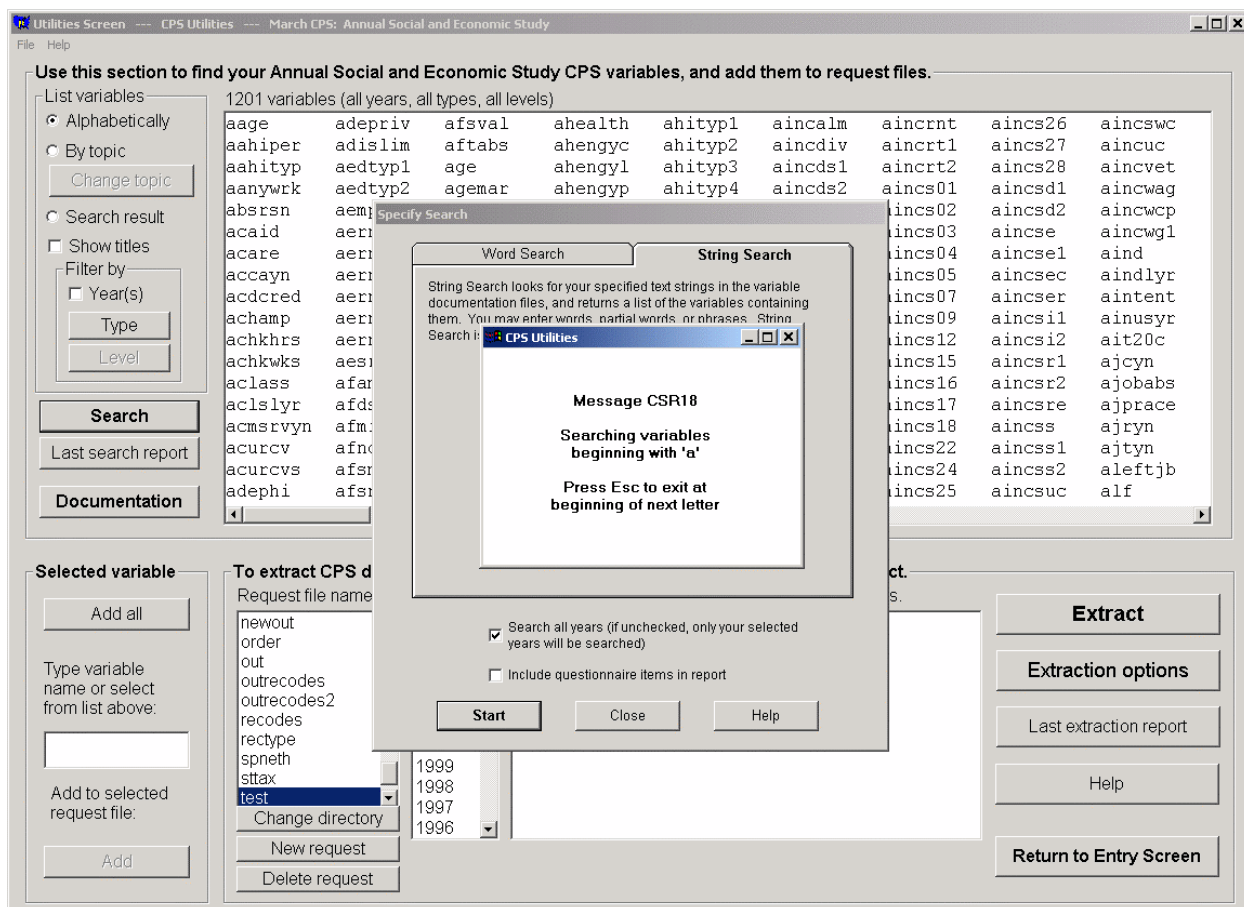


Figure 6. Specify Search screen for String Search

For example, to list documentation for all variables that relate to workers, you could enter the word 'worker' in the top box. Choosing Start initiates a search for the presence of the character string 'worker'. Each file in the Data Dictionary for your selected survey is searched for the presence of the given character strings. A progress window is displayed, showing the beginning letter of the variables currently being searched.

Because this SEARCH utility searches mechanically for the given character strings, it may produce unwanted dictionary entries. For example, the search for age-related variables initiated by searching on the word 'age' will also locate all entries containing the word 'wage'. In addition, the function is case-sensitive: searching on 'Span' will locate all entries containing the word

Spanish, but a search on 'span' will not.

Searches for multiple strings (up to five) may be requested with one SEARCH command. Option buttons allow the user to choose whether to flag entries containing any (the union) of the strings, or all (the intersection) of the strings. In either case, the variables flagged are displayed in alphabetical order, not in the order flagged. For example, if the union option is selected, and the words 'pension' and 'business' are entered, the variables whose dictionary entries contain the string 'pension' will be intermingled with the variables whose dictionary entries contain the string 'business', and all of the entries will be included in the results.

Other options available with this utility may be selected by checking the respective checkboxes. If "Include questionnaire items in report" is checked, the questionnaire items that led to the variable are included with the other dictionary data; otherwise they are not. We emphasize though, that the search for character strings does cover the questionnaire items as well. Thus, if the character string that you supply appears only in the questionnaire section, and you do not exercise the option to include questionnaire items in the report produced by SEARCH, you will not see evidence of the character string in the report.

The other option, "Limit search to topic description only" confines the string search to the single line of documentation that identifies the topic grouping in Section III of the data dictionary. That is, this more restrictive search looks for the user-supplied character string only in the topic to which each variable has been assigned. Note that it is rare to use this option with the intersection option because there is little substantive overlap in topic headings.

5.1.3 SEARCH RESULTS

The results of your search are displayed in the upper list on the Utilities screen.

The full dictionary entries for all variables on the list are saved to a file named cpssrch.rpt in your working directory. The contents of this file may be displayed on the screen by choosing "Last search report", and printed by choosing "Print report" or Print from the File menu. By examining this information, you may determine which variables match the concepts that you have in mind.

Be forewarned that if cpssrch.rpt already exists in your working directory it gets overwritten. The report file is not intended to be anything more than a temporary file to assist in identifying possible variables to include in a subsequent EXTRACT request.

For more help, press F1 or choose Help from the menu bar.

5.1.4. SAMPLE OUTPUT FILES: SEARCH UTILITY

To see the kind of documentation provided through String Search, consider the following results that are contained in the cpssrch.rpt produced by entering the word "ethnic" in the top box on the "String Search" screen.

CPS Utilities String Search Report

March CPS Dictionary
Annual Demographic and Income Sources

Search keyword(s): ethnic
Search year(s): All years in survey

Search specifications:
Union option applied
Questionnaire items not printed
Keyword(s) search done

aspneth Allocation flag: spneth

Original location, length, and name:
02-88B
P499
1
A%ORIGIN

Type: basic

Topic: demography item allocation flag

Related variable:
spneth - spanish ethnicity

Code:

	02-88B
No change	0
Blank to value	2
Value to value	3
Allocated	4
Value to value - no error	5
Blank to not allocated code	8

Note: Universe - all persons

ethnic Ethnic origin

Original location, length, and name:
75-71
P197
2

Type: basic

Topic: demography

Code:

	75-71
German	01
Italian	02
Irish	03

French	04
Polish	05
Russian	06
English, Scot, Welsh	07
Mexicano, Chicano	08
Puerto Rican	09
Cuban	10
Central or South American	11
Other Spanish	12
Negro	13
Other	14
Don't know	15
Not allocated	16

Note: Universe - all persons

hdspan Family head of Spanish origin

Original location, length, and name:

05-88B 88-76 75-71

F231 F19 F146

1 1 1

FSPANISH FSPNHEAD

Type: supplement

Topic: demography (household)

Code:

	05-71
Head is Spanish	1
Head is not Spanish	2
Head NA	3

Note: Universe - all family records

spneth Spanish ethnicity
 _spneth Unicon recode - Spanish ethnicity

Original location, length, and name:

05-03 02-88B 88-76

P28 P27 P113

1 2 2

PRDTHSP A-REORGN ETHNCITY

Type: basic

Topic: demography

Related variable:

 aspneth - allocation flag

Code:

Recode:

	05-03	02-88B	88-76	05-76
Mexican American		01	10	1
Chicano		02	11	2
Mexican	1	03	12	3
Mexicano	1	03	13	3
Puerto Rican	2	04	14	4
Cuban	3	05	15	5
Central/South American	4	06	16	6
Other Spanish	5	07	17	7
Another not listed	0	08	30	8
Don't know		09	39	9
Not available		10	40	10

In 1994 forward this is an edited variable with additional possible values of: -1 (blank) -2 (don't know) -3 (refused) -9 (no response)

Note:

2005-2003 Universe - all persons identified as being Hispanic (hisp=1)

2002-1976 Universe - all interview records (1<=popstat<=3)

CAUTION - The 'select' and 'unselect' options should be used carefully with the UNICON RECODE, as selection occurs after recoding of the variable. (i.e., The options use the recoded values rather than the original raw values.)

_spneth Spanish ethnicity - Unicon recode

Original location, length, and name:

05-03	02-88B	88-76	75-71
P28	P27	P113	P197
1	2	2	2
PRDTHSP	A-REORGN	ETHNCITY	

Type: basic

Topic: demography

Code:	05-03	02-88B	88-76	75-71	Recode:
					05-71
Mexican American		01	10	08	01
Chicano		02	11		02
Mexican	01	03	12		03
Mexicano	01	03	13		03
Puerto Rican	02	04	14	09	04
Cuban	03	05	15	10	05
Central/South American	04	06	16	11	06
Other Spanish	05	07	17	12	07
Another not listed	00	08	30	1-7,13,14	08
Don't know		09	39	15	09
Not available		10	40	16	10

Note:

2005-2003 Universe - all persons identified as being Hispanic

(hisp=1)
2002-1971 Universe - all interview records (1<=popstat<=3)

Prior to 1983, Hispanic populations in the CPS were not controlled to independent totals so the counts are not as reliable as the counts in 1983 and later. (per Census 9/03)

CAUTION - The conditional recode option does not operate on this variable. Using the '_' variable will always result in the default recoding.

CAUTION - The 'select' and 'unselect' options should be used Carefully with this UNICON RECODE as selection occurs after recoding of the variable. (i.e., The options use the recoded values rather than the original raw values.)

Total number of matching entries is 5.

Much of the documentation provided is self-explanatory, but some is not. We may use the first dictionary entry to illustrate. On the first line of the documentation is the name of variable (i.e., "aspneth"), followed by a brief description (i.e., "Allocation flag: spneth"). The next line shows the CPS year group(s) (i.e., 02-88B). If the variable is present in the indicated year group, the subsequent line gives the record type (i.e., "P" for personal record, as opposed to "F" for family records and "H" for household records) and column location (i.e., column 499). Immediately below the record type and column location is the length of the variable in columns (i.e., 1). Next is the type of the variable (i.e., "basic") and topic under which we have grouped the variable in the topical index (i.e., "demography item allocation flag"). Below the topic heading is a list of all codes and their meanings. Information on the universe of respondents follows (i.e., "all persons").

In the third entry, the string "ethnic" does not appear in the report. One may conclude that this string appears in the questionnaire text, which was not printed in the report because the "Include questionnaire items in report" option was not selected.

The last of the five entries is a recoded variable "_spneth". The documentation provides the codes for the original, inconsistently coded variable ("spneth"), as well as for the recoded variable that features consistent coding across the years. As noted, the Stata and SAS program code necessary to perform this recode will be provided upon executing EXTRACT, as described below. The cautionary note that is part of the documentation for "_spneth" is related to an EXTRACT option that will also be discussed later.

5.2. DOCUMENTATION UTILITY

If you already know (or can guess) the name of a variable of interest, you can display its documentation immediately with this utility.

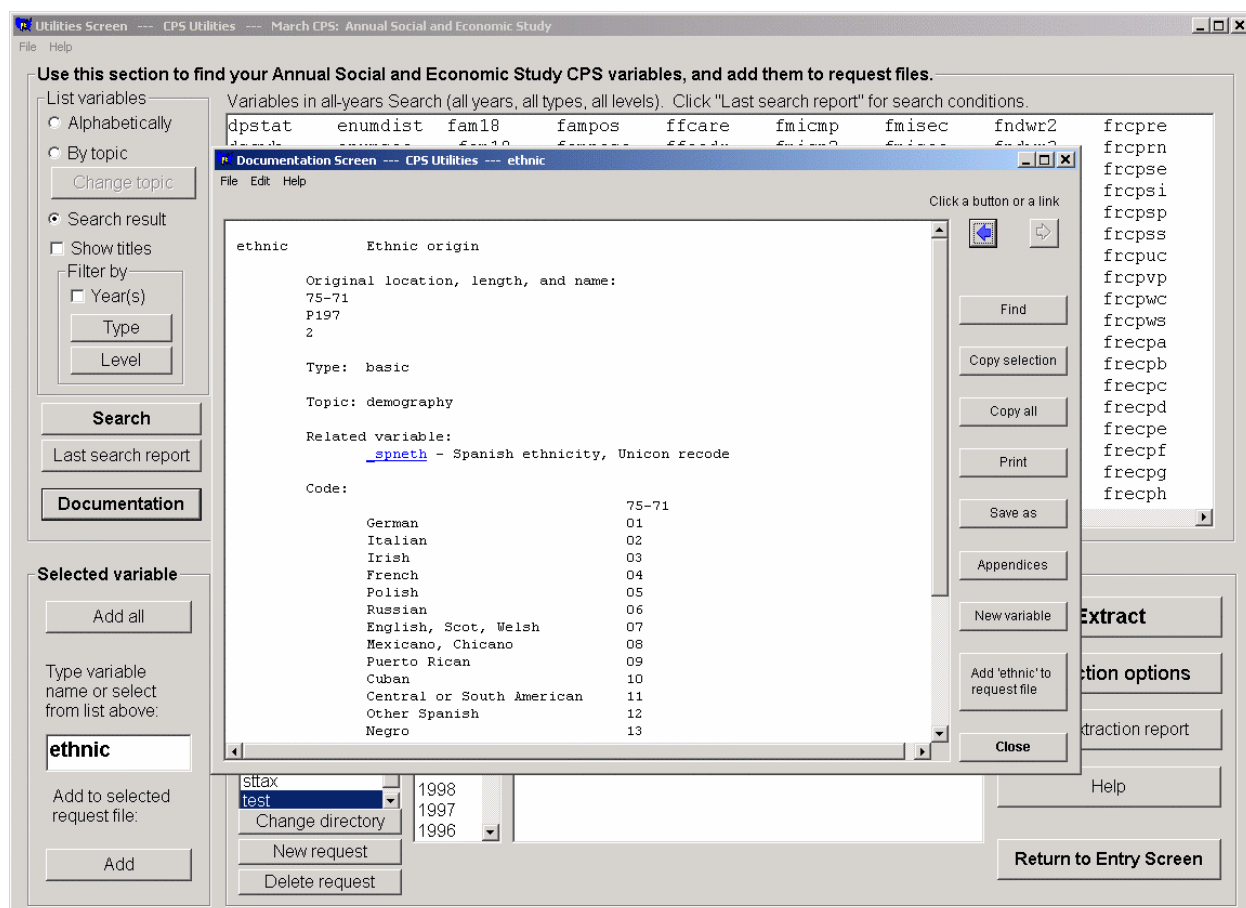


Figure 7. Documentation screen

Select your variable of interest from the upper list, then click 'Documentation'. The documentation details are displayed in a browser-like screen (Figure 7). 'Find' lets you search the current page for text. 'Copy' buttons are provided to let you copy selected, or all, text to the clipboard, which can then be copied into an editor such as Notepad. 'Print' lets you print the current page. 'Save as' lets you save the current page as either an .htm or .txt file. You can also use 'Appendices' to view information in an appendix. If you wish to look at documentation for another variable, click 'New variable' and select from the list that is displayed on a new window. Use the arrow buttons to move backward and forward through the pages of documentation variables. Click 'Add ... to request file' if the currently displayed variable matches your extraction requirements.

You can also access this utility by double-clicking on a variable in either the upper or lower list, or by selecting "View documentation ..." on the edit options menu. See Section 5.3.3 of this manual for details. For more help, press F1 or choose Help from the menu bar.

5.2.1. SAMPLE OUTPUT: DOCUMENTATION UTILITY

As an example, suppose that we wish to see the documentation for items that led to the variable "ethnic", which we discussed above. Selecting the variable "ethnic" from the list on the DOCUMENTATION screen yields the following information:

```
ethnic          Ethnic origin

Original location, length, and name:
75-71
P197
2

Type:  basic

Topic: demography

Code:
                                75-71
    German                      01
    Italian                     02
    Irish                       03
    French                      04
    Polish                      05
    Russian                     06
    English, Scot, Welsh        07
    Mexicano, Chicano           08
    Puerto Rican                09
    Cuban                       10
    Central or South American   11
    Other Spanish               12
    Negro                       13
    Other                       14
    Don't know                   15
    Not allocated                16

Note:  Universe - all persons

** Survey Questionnaire Source Items - ethnic **
    1975                33. Origin.
    1974-1971           52. What is... 's origin or descent?
```

Note that this utility always includes the questionnaire source items in the retrieved documentation. In this case it shows the question number (33) and item that led to this variable in 1975, and the question number (52) and question wording that generated this variable in the years 1971 through 1974.

Sometimes a variable is derived from other variables. For example, the variable "hdspan", for which documentation is presented in the report above, is derived from two other variables, and produces the following:

hdspan Family head of Spanish origin

Original location, length, and name:

05-88B 88-76 75-71

F231 F19 F146

1 1 1

FSPANISH FSPNHEAD

Type: supplement

Topic: demography (household)

Code:

	05-71
Head is Spanish	1
Head is not Spanish	2
Head NA	3

Note: Universe - all family records

** Survey Questionnaire Source Items - hdspan **

2005-1976 Derived from spneth, relhd

1975-1971 Derived from ethnic, relhd

Other variables, such as sample weights, are derived by the Census Bureau using information that goes well beyond the content of the CPS survey. The questionnaire documentation listed by DOCUMENTATION may refer to such variables as "office items."

5.3 EXTRACT UTILITY

The third utility, and indeed the heart of the whole CPS Utilities system, is EXTRACT. From a list of variables that you supply, this utility produces an ASCII data file containing the raw data from the Census Bureau data files; an extraction report containing the variable definition, coding, and universe information from the Data Dictionary; and optionally segments of program code to facilitate using the data in Stata, SAS, or SPSS. In the event that you do not utilize these statistical analysis programs, we believe that you can modify the code relatively easily to produce code in an alternative software package of your choosing. If Stata is installed on your system, you may start and load your extracted data with one button click.

Your list of variables to be extracted must be prepared as a request file prior to running an extraction. See Section 5.3.3 for full details of request file formatting and the facilities provided by CPS Utilities for creating and editing request files.

The raw data file produced by Extract may be an ASCII file containing variable values delimited by newline characters, i.e. a single vertical column of values. Each record (observation) is represented by a contiguous list of values, of length determined by the number of variables selected by the user in the request file (see Section 7.3.3), usually one value per variable. The values for successive records follow with no separate record delimiter. The sequence of variable values within a record is indicated in the Stata, SAS, and SPSS dataset creation program code files, and matches the sequence in the report file. Optionally, the raw file may be produced in a

tabular format, with values of variables within a record delimited by spaces or commas, and records delimited by newlines; variable names may also, optionally, be included as the first record in the table. A third table option is to right justify the values in fixed-width fields, with each record appended to the previous one without newlines. This option also creates a dictionary file, which contains the variable name, field width and column positions within the record. A further option enables a Stata dataset to be produced directly in lieu of the raw data file and the dataset creation program segments that would input the raw data.

Recoded values are created dynamically and output directly in the raw data file or Stata dataset. These include more than eighty March variables which are created by CPS Utilities (identified by a leading underscore), and a number of Outgoing Rotation variables which change in mid-year and therefore require recoding as a function of the 'month' variable.

Missing values, if not specifically coded as missing in the Census Bureau files, are shown in the CPS Utilities output files as a dot. A dot may represent either an entry consisting of nothing but spaces, dots, and dashes, an Outgoing Rotation variable that exists in the current year but not in the month of the current record, or a variable that does not exist for the record's interview type.

Since 1994, the Census Bureau has used a new set of missing values for all unedited and some edited variables. The values are:

-1	Blank
-2	Don't know
-3	Refused
-5 or -9	No Response

The extraction report includes a header section indicating the date and time of the extraction, the survey year, and the names and locations of the request file, the input raw data file, and the output files. Also in the header section are the counts of records read and written, the total number of records available to be written, a table summarizing the values of each requested variable, and a list of options selected. Then follows, optionally, a list of attributes of the requested input file, and one or more tables showing attributes of the requested variables.

In addition, the extraction report includes copies of the documentation files for each of the selected variables. If a range of years is selected, this information is only included in the report for the first year to be extracted. If you need this information for variables which do not exist that year, you may run a separate "preview" report for a year in which the variable did exist. See the "Preview report only" option, below. If the "Combine years" option is selected, there is only one report so this information is provided for all variables included in the selected years.

If recoded variables are included in the request file, and the "Conditional recoding" option is selected (see below), this report also contains a section documenting the recoded values applicable to the user's specific extraction.

The report may be printed by choosing "Print report" or Print from the File menu

The dataset creation program code segments may be used to read the raw data file into Stata, SAS, or SPSS in the correct sequence, and to add value labels when appropriate, and variable

labels. The 'Execute code' button is provided to automatically start Stata, SAS or SPSS, and to load the program code. If you selected the comma delimited table option, 'Execute code' will automatically load the .csv file into a spreadsheet program, like Excel. For more information, choose 'Help' and then 'Help for this Screen'.

Please refer to Section 6 of this manual for more information about the structure of the CPS Utilities data files, and for tips on using the Utilities.

5.3.1 USING THE EXTRACT UTILITY

The Utilities screen (Figure 3 at the beginning of this Section) allows the user to select a request file, one or more years, and other options, described below. For the Monthly Basic survey, the available months are shown instead of years. On choosing EXTRACT, output files are produced, containing data on the variables listed in the request file, for the selected survey and year(s) (or month(s)).

If there are no request files in the working directory, a screen message prompts the user to make one. To make a request file, choose the "New request" button, and refer to Section 5.3.3.

The files produced by EXTRACT are written to a sub-directory of the working directory named for the request file name. If it already exists, any files previously created for the same period of data are overwritten. If it does not already exist, EXTRACT creates it.

EXTRACT's output files are named for the survey and year (or year and month for the Monthly Basic survey), with an extension indicating the type of file. For instance, mar05.rpt is the extraction report produced by EXTRACT when the user selects the Annual Social and Economic (March) survey, and the year 2005. Other files produced, depending on options selected, may be:

mar05.raw	Raw data file in single column, space-delimited or fixed width field format
mar05.csv	Raw data file in comma delimited format
mar05.do	Stata "do" file
mar05.dct	Stata dictionary file
mar05.dta	Stata dataset
mar05.sas	SAS code
mar05.sps	SPSS code
mar05.fxd	Fixed width field dictionary file

Monthly Basic output files are named mon9801.raw, etc. Combined year files are named for the survey plus "mult", as in marmult.raw, monmult.raw, etc.

On completion of an extraction, the Results screen (Figure 8) shows, initially showing the latest extraction report. If multiple years were selected and not combined, the 'Display other years' list will show, allowing the reports of the extractions from the other years to be displayed.

Option buttons under 'Show last results' allow any of the other files produced by this extraction to be displayed. In addition, the 'Tabulated data' button displays a table of the extracted data for visual inspection. Each column of the table is headed by the variable name, hyperlinked to its documentation file.

Other buttons on the Results screen allow you to find, copy, or print, as described earlier for the Documentation screen.

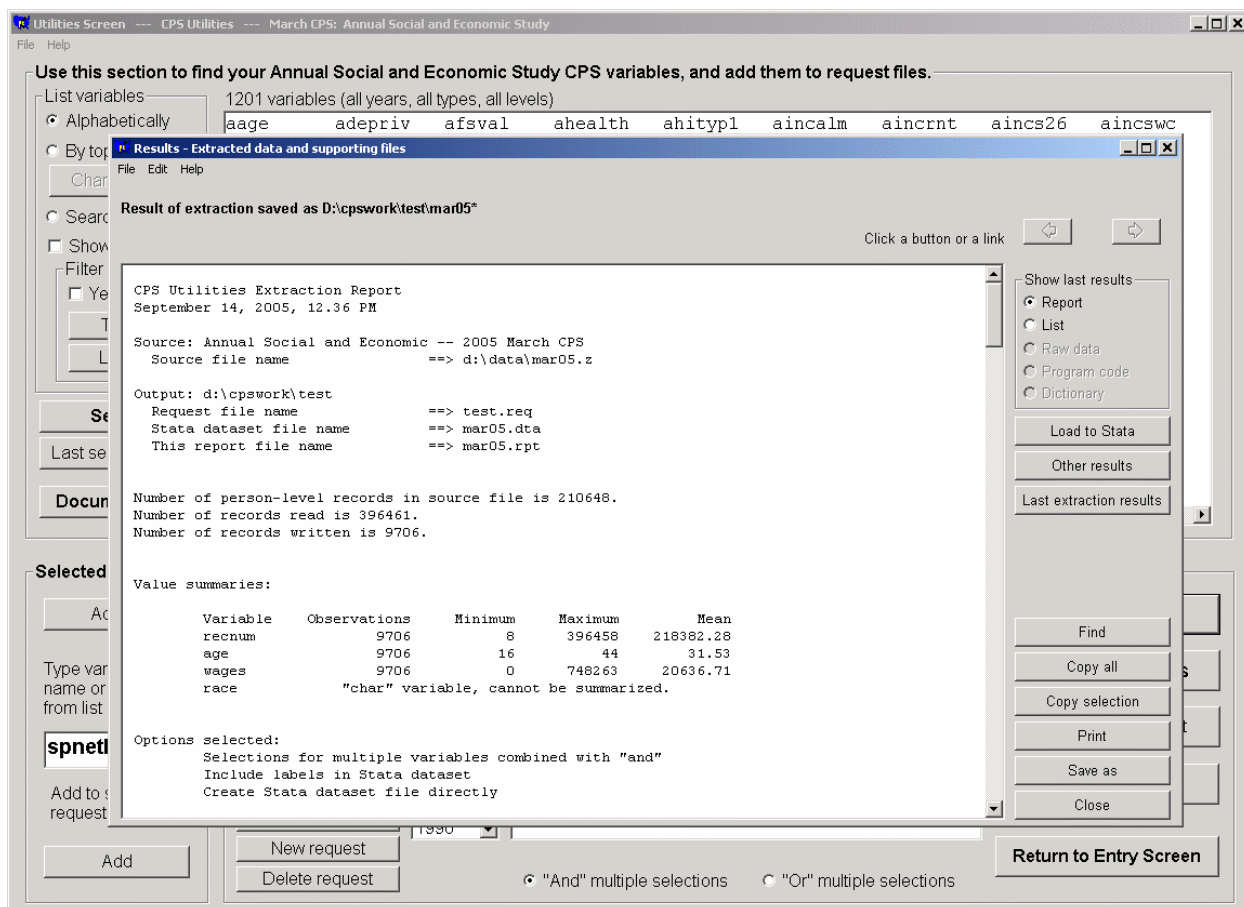


Figure 8. Results screen, showing Extraction Report

Except for the Stata dataset, which is not a text file, the contents of these files can be displayed on the screen at the completion of the extraction by clicking the appropriate file option. Also displayed is the full path to the file's location.

If a program loading code format is selected, the code may be displayed on the Results screen by selecting the applicable button under 'Show last results'. In this case, the raw data file can also be displayed in a similar way. These dataset creation program code segments may be used to

read the raw data file into Stata, SAS, or SPSS in the correct sequence, and to add value labels when appropriate, and variable labels. The loading button is labeled 'Execute code'. If Stata loading code was selected, it automatically starts Stata, and loads it with the raw data using the program code. For SAS and SPSS it starts the respective code editor, if installed.

The 'Execute code' button is labeled differently and has a different, but equivalent, effect if other output data formats are selected. For instance, if a comma-delimited format is specified, it is labeled 'Load to spreadsheet' and when clicked it opens Excel and loads the extracted data into it. Or if a Stata dataset is specified, it is labeled 'Load to Stata'. When this button is clicked, if Stata is installed on your computer, Stata is started and the current dataset is loaded into it. If the 'Select other file' list is showing, other datasets may be selected and loaded to Stata in the same way.

These files are intended for use on only a temporary basis. If you wish to keep them for long-term use, you may want to rename them. The risk in not renaming the files is that if you use the same request file name when you ask for other variables from the same survey (March 2002 in our example), the existing files with the same names are overwritten.

A CPS Utilities extraction may normally be aborted by pressing the Esc key, as indicated in the status message displayed during an extraction. However, if other keys are pressed before the Esc key, they may render the Esc key ineffective. The user must then wait until the current year's extraction is complete, when a screen message allows the extraction to be aborted before the next year in the selected range is started.

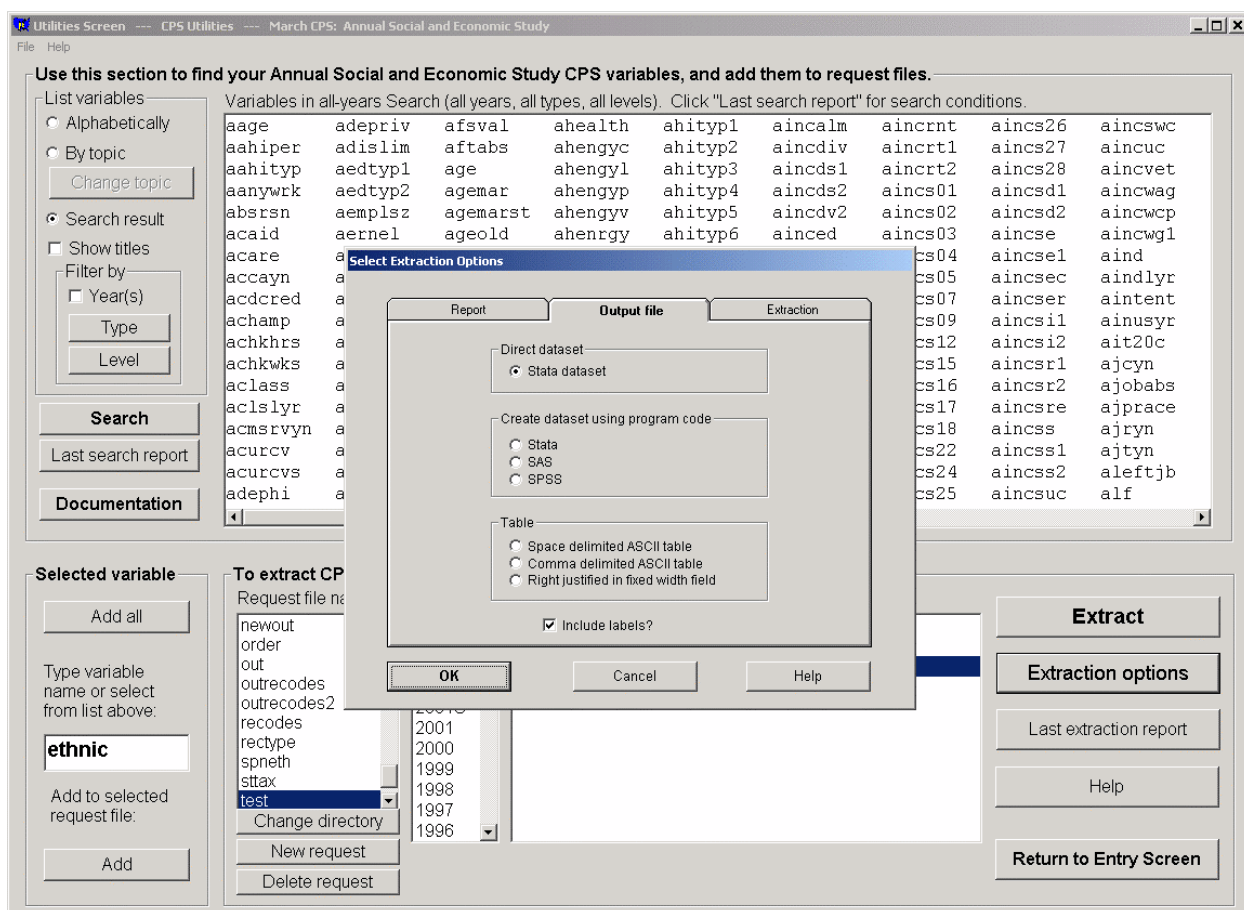


Figure 9. Extraction Options screen

The options available (Figure 9) with this utility may be selected by clicking "Extraction options". A window with four tabs is displayed. Click the appropriate tab, then applicable option buttons or check boxes, as follows:

"Report" tab

This allows you to choose which of these options you want to see in your report. You may choose any combination from none to all four.

Alphasort variables in report: This option places the descriptions of the requested variables in the extraction report in alphabetical order. The default is to place them in the order listed in the request file. Renamed variables appear in the order dictated by the new name, not the original name. This option also alphasorts the conditional recoding information, if that option is selected on the "Extraction" tab.

Include Questionnaire items in report: This option includes the survey source questions for each requested variable in the extraction report.

Omit data for unselected years from report: This option causes the extraction report to drop references to attributes of variables in years earlier than the beginning of the

range of years for which extraction is requested, and also omits conditional recoding information for years not included in the user-specified conditional recode range.

Include input file attributes in report: This option places input file attributes (column position, length, selection criteria, etc.) in the extraction report. Including them was standard in prior versions of CPS Utilities, optional in this version.

"Output file" tab

This allows you to choose one output file format. In all cases, the report file (.rpt) is created. A check box on the bottom of the tab lets you include labels for direct or created datasets (value labels (SAS formats) (when appropriate) and variable labels in the program code for Stata, SAS and SPSS, and in the Stata dataset) or to include the variable names as the first record in space or comma delimited tables.

Direct dataset

Stata dataset: This option allows the user to create a Stata dataset file (.dta) in one step. This saves disk space and time. The dataset may be loaded directly into Stata by choosing the 'Load to Stata' button.

Create dataset using program code

Stata, SAS or SPSS: This allows you to choose which dataset creation program code output you require. In all cases, an ASCII file (.raw) containing a single vertical column of values delimited by newline characters is output, as well as code files for SAS (.sas), SPSS (.sps), or Stata (.do) and a Stata dictionary file (.dct). The dataset may be loaded into your statistical program by choosing the 'Execute code' button.

Table

Space delimited ASCII table: This option produces the file (.raw) in a tabular format, with values of variables within a record delimited by spaces, and successive records delimited by newlines. This format is easier to read visually, and to enter into a spreadsheet, but cannot be used with the program code segments produced by CPS Utilities to load the file into Stata, SAS, or SPSS.

Comma delimited ASCII table: This option is the same as the space delimited format above, except the values of variables within a record are delimited by commas and the file name extension is .csv. The table may be loaded into your spreadsheet program by choosing the 'Execute code' button.

Right justified in fixed width field: This option right justifies, with leading spaces, values in a fixed width field. The fields have no delimiters within a record, and the records have no delimiters within the file. A dictionary file (.fxd) is created, showing the variable name, field width and column positions within a record. If a variable is "asis", it will be output exactly as it is in the source data file, including leading and/or trailing spaces or leading zeroes, rather than changed to right justified.

"Extraction" tab

This allows you to choose which of these options you want to include in your extraction.

You may choose any combination from none to all five.

Include labels: This option places value labels (SAS formats) (when appropriate) and variable labels in the program code for Stata, SAS and SPSS, and in the Stata dataset.

Combine years: This option creates a single output file and report containing the combined data from all the years selected for the extraction. It is available for all output formats. The variables are output in the same order as they are in the request file regardless of year, with missing data represented by a dot.

Preview report only: This option produces the available sections of the extraction report of the selected variables without running the extraction. This allows the user to peruse the documentation of the proposed file and make any changes before committing the time and resources to run the job.

Conditional recoding: This option allows the user to select a range of years over which uniform recoding is required. If this option is not selected, default recoding is used, which provides uniform recoding over all survey years, possibly resulting in unnecessary information loss. See Section 5.3.5 of this manual, or the Windows on-screen Help messages, for further information.

Limit number of records written: This option restricts the number of observations written to the output files to be no greater than the number entered in the box below. Due to the hierarchical format of the March files, and any selections the user may have made on the values of the variables to be extracted, the number of records read in may be more than that written out.

For more help, press F1 or choose Help from the menu bar.

5.3.2 MULTITASKING

Users may run other applications while an extraction is running, as follows.

After setting the options and clicking Start, some screen messages may appear, which require user response. If a variable is not available for a year there will be a warning message. If you already have extraction files for this request file and year, there will be a message asking if you want to overwrite the existing files. Once the user has responded to these possible messages, the extraction continues unattended to completion, or until an error condition is encountered.

Other open applications may then be utilized by clicking them on the taskbar. Clicking the Start button on the taskbar and using the Start Menu may start new applications.

Open applications may be accessed from the keyboard with Alt-Tab, and new applications may be started with Ctrl-Esc. It is not possible for the user to minimize, or resize the CPS Utilities window while an extraction is running.

When the CPS Utilities extraction is completed, or if an error occurs, the CPS Utilities window

returns. At this point the user may process the results of the extraction, or minimize, move, or resize the CPS Utilities window to finish work in the other application.

5.3.3 REQUEST FILES

The request file, "test.req" in our example, is merely a list of desired variables, one per line. This file must not include any blank lines, for EXTRACT will attempt to read blank lines as variable names. The variables must match names given in the Data Dictionary. Comments may be included in the request file if the first character(s) are "*" or "/" or if surrounded by "/* ... */".

On choosing "New request" from the Utilities screen, a prompt asking the user for a file name is displayed. A new file may be created as a normal empty file, by copying an existing file, or combining two existing files and eliminating duplicates. . File names may be longer than eight characters and may contain spaces.

Note: If you use spaces in your file name and use Version 7 of Stata, you may encounter difficulty loading your datasets. Download the latest update of Stata 7 from their website, <http://www.stata.com>, and that difficulty will go away.

If the new file was created as an empty file, the 'recnum' variable is automatically added. Variables are displayed in the upper list. Clicking "Alphabetically" displays all variables for the survey. If you wish to view variables within a topic, click "By topic" or "Change topic" to view a different topic. "Filter by year" shows only those variables that are available for the selected years. "Filter by type" shows only those variables whose origin are the supplement or basic questionnaires or generated internally by the Census. "Filter by level" is available for the March Annual Social and Economic Study and Child surveys only, and shows only those variables within household, family and person hierarchical levels. If "Show titles" is checked, the variable names and a brief description are displayed in one vertical list. If it is not checked, just the variable names are displayed in eight vertical columns.

If by using a combination of topic and year selections the upper list contains all the variables you wish to extract, click "Add all" to copy all the variables from the upper list into your request file contents box. Variables may also be added through other means: highlight a variable on the upper list and drag it to the contents box; highlight it and click "Add"; or type the name in the Selected variable box and click "Add". Similarly, multiple variables may be selected using Ctrl and/or Shift keys. If the key remains pressed, all selected variables can be dragged to the contents box; or notice the Selected variable box shows the term '(multiple)', and just click "Add". In our example, the variables 'age', 'incwag' and 'race' have been added to the request file.

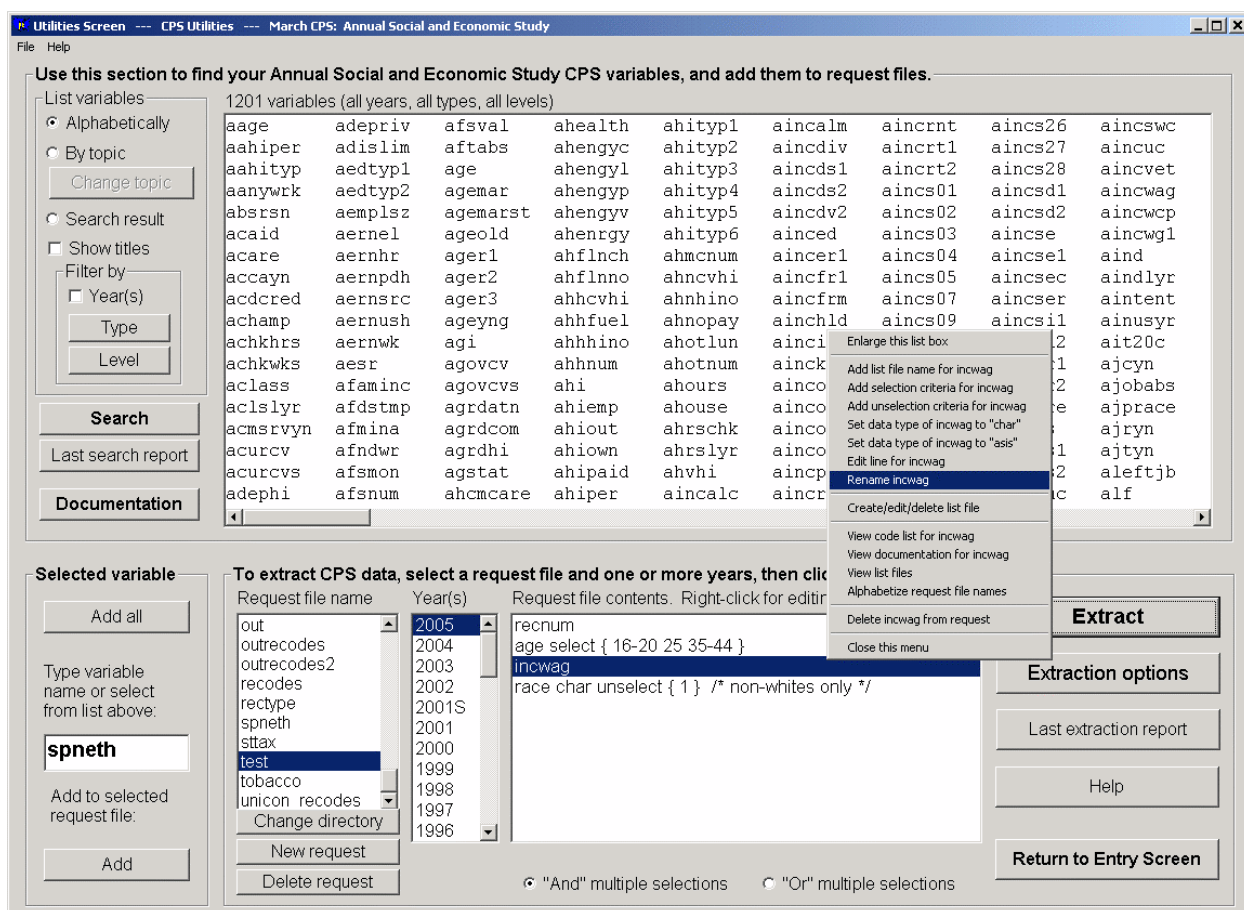


Figure 10. Edit Options menu

If desired, data type, value selection, and renaming options may be invoked within the request file itself by right-clicking the variable in the contents list and selecting a menu item. The "Edit options" menu (Figure 10) is displayed, offering many features to facilitate the preparation of request files meeting the required syntax criteria. The "... " below indicates where the actual variable name is used in the menu items.

Add list file name for ...: This is used to select the name of a list file that contains selection criteria for the selected variable. A window shows all files in your working directory which have the extension ".lst". Select a file and choose 'View contents' to see its contents in another window. Choose 'Add to ...' to attach the file to the variable in the contents box. The format for each item in a list file is the same as that for each entry in a selection criteria specification, as explained below. List file names may not be longer than eight characters, and may not contain spaces.

FOR MAY SURVEY ONLY: Add variable name range for ...: This is a special feature, which may be used to define a range of variables, which are similar, such as *course01* through *course23*, *hmequip1* through *hmequip6*, etc. It may be used with "select", "unselect" and "list", and assigns the same selection criteria to all of the variables in the range in just a couple of steps, instead of having to add each variable and

selection criteria individually. Suppose you want to extract variables *course01* through *course03* using selection criteria contained in list file "*class.lst*". Add *course01* to the request file, then use 'Add list file name' to add "class" to *course01* – so far nothing different. Now, select *course01* and choose 'Add variable name range for course01'. A list displays *course02* through *course23*; select *course03* and click 'Add to course01'. The request file now contains 'course01-03 list { class }', and the selection list is attached to all three variables automatically.

Add selection criteria for ...: This is used to specify criteria to limit the extracted data to records in which the selected variable has specific values. Note that you must enter numerical limits in the format shown in the screen instructions, or you may get error messages or incorrect results. If the data type is set to char or asis you may also select on alphanumeric strings.

Add unselect criteria for ...: This is used to specify criteria to limit the extracted data to records in which the selected variable does not have specific values. The same notes apply as for selection criteria.

Set data type of ... to "asis": This results in the values of the selected variable in the resulting raw file being exact copies of the text as published in the government files, enclosed in double quotes. Corresponding changes are made to the documentation report, the program code, and the Stata dataset. (In cases where the variable is absent, the value is represented by a dot, enclosed in double quotes).

Set data type of ... to "char": This has the same effect as the asis option, except that leading zeros and leading and trailing spaces are stripped, and values consisting of nothing but spaces, dots or dashes are converted to a dot.

Edit line for ...: Modify a line from the request file. Use this very carefully.

Rename ...: Specify a different name for the selected variable. The new name will be used in the output report file, the Stata, SAS and SPSS program code, and the Stata dataset. Names must not exceed eight characters in length, must begin with a letter or an underscore, and must contain only letters, numerals, or underscores.

View code list for ...: Displays the coding information in the documentation for the selected variable. If coding does not exist, a message is displayed and gives you the opportunity to invoke the Documentation utility.

View documentation for ...: Invokes the DOCUMENTATION utility for the selected variable.

Delete ... from request: Remove the selected variable from the request file.

Other items on the editing options menu are:

Enlarge this list box: If you want to see more variables in your request file than will

display in the contents box, click on this option to make a larger box. Right-click on a variable to see all the options for criteria setting as above.

Create/edit/delete list file: Lets you create list files, edit their contents using Notepad or delete them from your system.

View list files: A window shows all files in your working directory, which have the extension ".lst". Select a file and choose "View contents" to see its contents in another window. Choose "Add to ..." to attach the file to the variable in the contents box. The format for each item in a list file is the same as that for each entry in a selection criteria specification, as explained below.

Alphabetize request file names: This option sorts the variables in the request file by the original name – renames are ignored. Sorting is done only when the option is selected.

To close the menu, click anywhere on the Utilities screen.

If a selected variable already has "asis", "char", "select", "unselect", or "rename" criteria specified, the menu item changes to allow you to edit and/or delete the criteria.

Wildcards may be used in either numerical or alphanumeric selection criteria. An asterisk (*) as the last character of a selection criterion selects all values beginning with the characters preceding the asterisk, e.g. 57* selects 57, 570, 57864, etc. Questions marks (?) anywhere in a selection criterion selects all values where all the other characters match, e.g. 57?8? selects 57080, 57287, etc., but not 5708. Both wildcards may be used if required, e.g. 57?8* selects 5708, 5738, 5798125, etc.

If you have specified select or unselect criteria for more than one variable, you may specify whether the selections should be combined with "and" or "or". For example, suppose you have selected age values 0-49 and sex value 1 (male). If you combine these selections with "and", your extraction will only include records for males aged 0-49. If you combine them with "or", you will get the records for all persons who are either male, aged 0-49, or both.

The "select" and "unselect" options are designed for use with variables having relatively few outcomes, such as categorical variables. In particular, in any request file there can be no more than 8192 characters in the list of all of the "select" and "unselect" values together. For example, the "select" and "unselect" options in the small request file below specify 16 two-character values of age and 1 one-character value of race, for a total of 33 characters. Taking the spaces between the listed values into account, the list length becomes 49 characters. If the "select" and "unselect" options are used only for categorical variables, or for variables, like age, which assume relatively few integer values, the limit of 8192 characters is not very restrictive.

On the other hand, using select on a continuous variable, such as incwag, may easily result in more than 8192 characters, so should be avoided. For example,

```
incwag select { 20000-29999 }
```

results in an error message and no extraction. Instead, use wildcards to get the required result:

```
incwag select { 2???? }
```

Caution must be exercised when using "select" or "unselect" options with recoded variables. These options operate on the recoded values of the CPS variables. If you use the Conditional Recode option, these values may vary depending on the range of years you select. To remind the user of this fact, the documentation contains a warning message like that shown for the recode "_spneth", in the SEARCH report example above.

For users who use a text editor to create their request files, the syntax for these five options is as follows:

- Follow the variable name by a space and one of the key words "asis", "char", "select", "unselect", or "rename".
- If "char" or "asis" are used together with any of the other keywords, "char" or "asis" must appear first, followed by a space, and the next key word.
- Follow the key words "select", "unselect", or "rename" by a space, an opening brace ({}), and another space.
- In the case of "rename", then enter the new name, beginning with a lower-case letter or an underscore (_), and including only lower-case letters, numerals and underscores. Then enter a space and a closing brace ({}).
- In the case of "select" or "unselect", enter any number of positive numerals or ranges of numerals, followed by a space and a closing brace. Show ranges as two numerals separated by a dash (-), with no spaces. If you have more than one numeral or range, separate them by spaces. Since dashes are used to denote ranges, it is not possible to specify negative values in the select and unselect options.
- You cannot use both "select" and "unselect" on the same variable, but you may use "rename" either before or after "select" or "unselect". Separate them by spaces.
- Extra spaces do not matter.

To illustrate, the example request file is:

```
recnum
age select { 16-20 25 35-44 }
incwag rename { wages }
race char unselect { 1 } /* non-whites only */
```

This request file specifies that the variables 'recnum', 'age', 'incwag', and 'race' are to be written to a file. Moreover, if the multiple selections are specified to be combined with "and", only cases for which 'age' is 16 through 20, 25, or 35 through 44, and 'race' does not have a value of 1, are to be included. The variable 'incwag' now appears in the report and the program code as 'wages'. The values of the variable race will be enclosed in double quotes in the raw file, and the program code will specify its data type as character. The comment enclosed between the symbols /* and */ is ignored.

For more help, press F1 or choose Help from the menu bar.

5.3.4 SAMPLE OUTPUT FILES: EXTRACT UTILITY

As examples of the SAS, SPSS, and Stata code that is produced by EXTRACT, consider the following program segments that were produced by an EXTRACT operation in which the variables from the example request file were read from the March 2005 file:

```
* Program code to read raw data into SAS. ;
* Please review it to be sure it works with your system. ;

filename raw 'd:\cpswork\test\mar05.raw' ;
libname path 'd:\cpswork\test' ;
options nocenter ;
data path.extract ;
infile raw dlm='"' ;

length recnum 8 ;
length age 3 ;
length wages 4 ;
length race $ 2 ;
input
    recnum /
    age /
    wages /
    race
;

* Program code to specify labels. ;

label recnum = "Unique record ID number" ;
label age = "Age of person" ;
label wages = "Income from wage & salary" ;
label race = "Race" ;

* Program code to specify formats. Edit if necessary if you
renamed any variables. ;
```

```

proc format ;
  value fage      /* age */  80="80-84" 85="85+"
  -1="blank" -2="dontknow" -3="refused" -4="hrs vary" -9="no resp" ;
  value frace /*race*/ 1 = "Wh" 2 = "Blk" 3 = "AI" 4="Asian"
  5="HP" 6="W+B" 7="W+AI" 8="W+A" 9="W+HP" 10="B+AI" 11="B+A"
  12="B+HP" 13="AI+A" 14="A+HP" 15="W+B+AI" 16="W+B+A" 17="W+AI+A"
  18="W+A+HP" 19="W+B+AI+A" 20="2-3races" 21="4-5races"
  -1="blank" -2="dontknow" -3="refused" -4="hrs vary" -9="no resp" ;

proc contents ;
run ;

```

```

* Program code to read raw data into SPSS .

* Note 1: Any variable names beginning with an underscore have been renamed
to begin with an at-sign (@), to comply with SPSS naming rules .

* Note 2: If you move the raw file from your working directory to another
location, you must change the path shown in the FILE= statement below .

SET
  UNDEFINED= WARN.
DATA LIST
  FILE='d:\cpsdata\test\mar05.raw' FREE /
recnum *
age *
wages *
race * .
EXECUTE .

* Program code to specify labels.

variable labels recnum "Unique record ID number" .
variable labels age "Age of person" .
variable labels wages "Income: wage & salary".
variable labels race "Race" .

```

```

* Program code to read raw data into Stata, using the
* accompanying .dct file.

* Note 1: you may save disk space by using the Stata dataset option,
* which generates a Stata .dta file directly.

#delimit ;
set more off ;
infile using mar05.dct ;

rename v1 recnum ;
rename v2 age ;
rename v3 wages ;
rename v4 race ;

* Program code to specify labels. ;

label var recnum "Unique record ID number" ;
label var age "Age of person" ;

```

```

cap label define fage 80 "80-84" 85 "85+"
-1 "blank" -2 "dontknow" -3 "refused" -4 "hrs vary" -9 "no resp" ;
label values age fage ;
label var wages "Income: wage & salary" ;
label var race "Race" ;
cap label define frace 1 "Wh" 2 "Blk" 3 "AI" 4 "Asian"
5 "HP" 6 "W+B" 7 "W+AI" 8 "W+A" 9 "W+HP" 10 "B+AI" 11 "B+A"
12 "B+HP" 13 "AI+A" 14 "A+HP" 15 "W+B+AI" 16 "W+B+A" 17 "W+AI+A"
18 "W+A+HP" 19 "W+B+AI+A" 20 "2-3races" 21 "4-5races"
-1 "blank" -2 "dontknow" -3 "refused" -4 "hrs vary" -9 "no resp" ;
label values race frace ;

```

```

dictionary using mar05.raw {
    long v1
    _newline
    byte v2
    _newline
    long v3
    _newline
    str2 v4
}

```

If you do not use Stata, SAS or SPSS, you may be able to use one of these program segments to get started. To compute the number of columns to allow for each variable that appears in the ASCII file (which contains only one variable per line), you may look up the maximum length of each variable in the codebook section of the Data Dictionary.

5.3.5 RECODES

The Data Dictionary for the March survey documents more than eighty Unicon-created recodes. These variables, which are denoted by variable names starting with an underscore "_", are of two types and serve two distinct purposes. One kind utilizes one or more raw CPS variables to generate a variable that is often used in empirical applications but is not available directly in the CPS. For example, the recode '_educ' (highest grade completed) is made from the raw variables 'grdhi' (highest grade attended) and 'grdcom' (whether highest grade was completed).

The other, more typical, recode is a uniformly coded version of a raw variable for which the coding changes over the years. In these circumstances, we offer a "least common denominator" recode.

That is, the recode contains only the information that can unambiguously be extracted from a range of years. For years in the range in which the raw variable features a more detailed coding, the recode lumps together categories, resulting in a corresponding loss of information.

The EXTRACT utility dynamically calculates the recoded values and outputs them in the raw data file or the Stata dataset.

The Conditional recoding option allows the user to specify the range of years over which uniform recoding is to take place. A smaller range reduces or eliminates the information loss.

If conditional recoding is not selected, CPS Utilities provides a default recoding giving uniformity over all survey years.

When the conditional recoding option is selected, a new screen appears, allowing the user to either select a uniform recoding range of years from a list, or cause the range to be the same as the range of years to be extracted.

Once this selection is made, it remains in effect until the user changes it by unchecking the conditional recoding check box on the "Extraction options|Extraction" screen. The caption on that check box changes to indicate whether the recoding range is set to a fixed custom setting, or will follow the extraction range.

It is very important to note that if the conditional recoding option is selected, the values of the variables do not necessarily represent the same information as they do when Conditional recoding is not selected. In fact the values may even change their meaning when the conditional recode range of years is changed.

Further, if you select the option on this screen to have the conditional recode range follow the extraction range, the values may change their meaning just because you changed the extraction range.

For these reasons, if you select the conditional recoding option, you must refer to the Conditional Recode information table in the extraction report (the .rpt file) to find which values have been recoded under your specific extraction conditions. Then you must refer to the documentation for each recoded variable, further down the report, to find what the values meant before they were recoded. (This part of the documentation also identifies the default recoding, which applies if you do not select conditional recoding).

For example, suppose you included the variable `_child18` in your request file, selected the conditional recoding option, and set your conditional recode range to 1965-1970. To assure uniform recoding across this range, CPS Utilities will recode any values of 8 or 9 appearing in the 1968-1970 data, to 7. This will occur whatever extraction range is selected. Assuming the "Omit data for unselected years from report" box was checked on the "Extraction options|Report" screen, the extraction report will include the following entry:

Conditional Recode information:

User-selected Conditional Recode range: 65-70

User-selected extraction range: 68-70

_child18, recoded from child18

In survey years 68-70:

8 recodes to 7 (Conditional Recode)

9 recodes to 7 (Conditional Recode)

10 recodes to .

Now refer to the coding information in the following section of the extraction report to find the meaning of 8 and 9 in years 1968 and later, and recognize that they had to be recoded to 7 to achieve uniformity with prior years:

Original location, length, and name:

05-88B 88-76 75-68 67-63

F27 F153 F63 F39

1 1 2 1

FOWNU18 FRCODE10

Type: basic

Topic: demography (family)

Code:

Recode:

	05-76	75-68	67-63	05-63
Number of children	0 - 9+	0 - 9+	0 - 7+	0 - 7+
NIU		10		.

Conditional Recode: If Conditional Recode option is selected, values depend on the selected years.

The default recode shown in this section of the documentation applies if the Conditional Recode option is not selected.

By contrast, if the conditional recode range is set to 1968-1970, there would be no need to recode the values 8 and 9, and the additional information represented by these values would be retained. In this case, the report simply states:

Conditional Recode information:

User-selected Conditional Recode range: 68-70

User-selected extraction range: 68-70

_child18, recoded from child18

In survey years 68-70:

10 recodes to .

The absence of any reference to values 8 and 9 indicates that they are unchanged under these conditions.

For this variable, the value 10 is recoded to a missing-value symbol (a dot) in all cases.

Your conditional recode range must include all the years in your extraction range. If you do not include them yourself, CPS Utilities will add them for you.

6. IMPORTANT INFORMATION AND TIPS FOR USERS

The following will help the new user understand the structure of the CPS data and the use of the CPS Utilities.

6.1 STRUCTURE OF THE CPS DATA FILES FOR MARCH AND CSA

The March CPS file for 1962 is a flat, rectangular file; i.e., it contains one record per person. After 1962, the March files in their original form are hierarchical; i.e., higher-level records contain data that are applicable to lower level records. From 1963 through 1975 the files have a Family-Person structure. That is, for each family, there is a family record, and under this record are the person records, one for each member of the family. From 1976 through the present, the March files have a structure of Household-Family-Person. That is, each household has a household record; under this record is a family record for each family (primary family and subfamilies). Under each family record are person records, one for each member of the family, beginning with the head of the household (for the primary family) or the head of the family (for subfamilies). The Child Support and Alimony series, which begins in 1979, is a match of the March and April files. Consequently, it too has the 3-tiered hierarchical structure.

The EXTRACT utility outputs ONLY person level records. If a user requests household or family variables, these values are retained and attached to each person record. If the user wants to make a file of only one record per household, it is necessary to write code in the user's chosen statistical application (e.g., SAS, Stata, SPSS, etc.) to keep only one record per household. This procedure should be easy to implement using the household ID variables ('hhid3' or 'hhid').

Notice in this regard that the documentation identifies the record level for each variable. Preceding the column location, there is a letter: 'H' for household level, 'F' for family level and 'P' for person level. Notice also that a family (household) level variable for an individual characterizes that person's family (household) as a whole, not just the individual. For example the variable 'child18' is a family level variable that indicates the number of children under the age of 18, never married, in the family. A young child may have a value up to 9 if there are several siblings in the family.

The "year" denoted as 1988B in the March survey is the 1988 Bridge file, which is the 1988 data, reformatted into 1989 variables and record layout. Comparing the results of the 1988 (original) file and the 1988 Bridge file gives users a sense of the questionnaire changes that were made between 1988 and 1989. There is a more detailed discussion of this file to be found in Appendix W of the Data Dictionary.

6.2 STRUCTURE OF THE CPS DATA FILES FOR OTHER THAN MARCH AND CSA

Files other than those for the March and CSA CPS are flat, rectangular files, not hierarchical. Hence, for CPS months other than March, there is only one record per person, and this record

contains all information relevant to that person.

Prior to 1994 the format of these records in the original data varied across interview types (adult civilian interviews, child interviews, armed forces interviews, or non-interviews (types A or B/C)), and the 'rectyp' variable was used to denote the interview type, and thus the format of the data. With software version 3.0 or later, the CPS Utilities is now able to access all record types. Most variables have consistent column locations across the different record types. However, some variables do have different locations in the adult record than in the child record. In these cases, the documentation shows two column locations (e.g., 101/187) where the first number is the location on the adult record and the second is the location on the child record. The universe description indicates the record types applicable for each variable. It is left to the user to select which records types to extract. To create a sample of only adult civilian interviews, the user should select only records for which 'rectyp' = 1. If non-interview and armed forces records are left in the sample, several key variables will appear to have a substantial number of records with missing data.

In 1994 and beyond, the same record layout is used for all interview types in the non-March files; there is no rectyp variable. To create a sample of adult civilian interviews, the user must select records for which 'intstat' = 1 and age is 15-99.

6.3 MISCELLANEOUS TIPS ON USING THE UTILITIES

We recommend that the 'recnum' variable be included in every request file. This variable, which is created by Unicon, not the Census Bureau, is simply a record number that is incremented as the data are read from the file (but prior to imposing any selection criteria). If, after extracting and modifying a file, the user realizes that more variables are needed, the user can make a second extraction from the SAME source file and use the 'recnum' variable to merge the two extractions. It is NOT meant to be used in merging files extracted from DIFFERENT source files. It is NOT meant to be used to match files across years.

In some instances, information for a variable may be missing after the Census collects the data. The Census may choose to allocate a new value for this missing datum, perhaps using a "hot deck" procedure. If the variable in question has allocated values, there is a corresponding allocation variable that indicates, in varying degrees of detail, whether and how the values were changed. This variable has the name of the original variable plus a prefix of 'a' for allocation. For example, the allocation variable for 'sex' is 'asex'. Appendix V in the Data Dictionary contains detailed information regarding data preparation and imputations used by the Census. Notice that the names applied to the variables have been assigned by Unicon. In the early years of the CPS data, there were no names on the variables in the Census documentation. In more recent years the Census documentation has supplied names for variables, but the names have tended to change over time. To take two examples of such changes, 'esr' (employment status recode) was changed to 'lfsr' (labor force status recode), and 'relhd' (relationship to head) was changed to 'rrp' (relationship to reference person). It is true that Census name changes tend to reflect evolution in the concepts relating to the variables (i.e., 'esr' actually classifies more than just employed people, but also people who are not employed but still in the labor force). Any such changes in Census definitions or concepts are indicated in the Unicon documentation.

Knowing that name changes can cause problems in existing computer programs, we attempt to retain the names that we originally assign. As we add more survey months, however, we make some changes in an effort to induce consistent naming across the various survey months. As a result, we may occasionally change variable names when we issue a new release of the data for a particular survey month. Any changes that are made across different releases of the same survey month are documented in Section II, 'Renamed Variables,' in the Data Dictionary.

Finally, Unicon is happy to field questions about the actual numbers that appear in a user's statistical results or about the concepts behind the various questionnaire items. We may, however, have to refer some questions to the CPS department of the Census Bureau (301 457-3806), for this is the department that puts the original files together. When calling the Census Bureau, the user must refer to variables by the column location and record type, NOT the name that Unicon has assigned to the variable.

7. FUTURE ENHANCEMENTS TO THE SOFTWARE

New releases of the CPS Utilities have been accompanied by additional options, some of which have been planned since the inception of the project, others of which have been suggested by users. We intend to add yet more options in the future, including the ability to select randomly subset of observations at a user-specified sampling rate.

We are open to all suggestions that will improve the speed or convenience of our system. If you wish to suggest changes or make comments, please feel to call us at Unicon (310 393-4636; ask for Eanswythe Grabowski or Bryan Rickard), or send us e-mail at support@unicon.com.

Also, be sure to check our website (<http://www.unicon.com>) periodically for announcements and updates.